REGIONAL DISPLACEMENT MATCHING SCHEME FOR LBP BASED FACE RECOGNITION

by

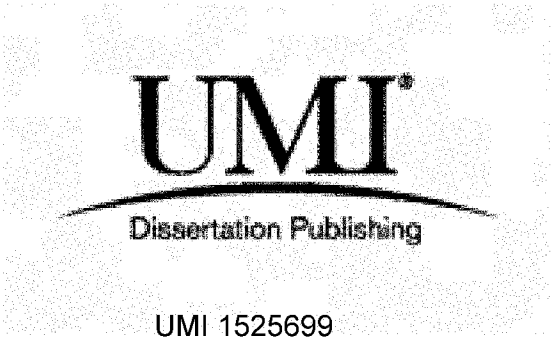
Ling Yan

B.Sc., Shandong University, 2005

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
MATHEMATICAL, COMPUTER, AND PHYSICAL SCIENCES
(COMPUTER SCIENCE)

THE UNIVERSITY OF NORTHERN BRITISH COLUMBIA

May, 2013



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

In face recognition, alignment of the face images has been a known open issue. This thesis proposes a displacement based local aligning scheme to construct a structural descriptive image template for comparison. To conquer the registration difficulties caused by the non-rigidity of human face images, a block displacement strategy is introduced to apply the regional voting scheme to face recognition field. Local Binary Pattern (LBP) is adopted to construct this block LBP displacement-based local matching approach, we name LBP-DLMA.

Experiments are performed and have demonstrated the outstanding performances of this LBP-DLMA over the original LBP approach. It is expected and shown by experiments that this approach applies to both large and small sized images, and that it also applies to descriptor approaches other than LBP.

Contents

List of Tables

List of Figures

Acknowledgement

Great thanks to my supervisor Dr. Liang Chen, who always has great confidence in me, who insists to name my playing with the data "experiment" and is always ready to provide his unreserved support. He is not only my academic adviser, but also a mentor and real friend (whose wife feeds me with the fancy foods that I have never had in my own kitchen).

Great thanks to my co-supervisor Dr. David Casperson, who generously squeezes me into his busy schedule all the time. I will benefit forever from his serious attitude in research. Dr. Casperson, merci beaucoup!

Thanks to my thesis committee member Dr. Jueyi Sui for his encouragement and insightful comments.

Thanks to my parents for their unconditional support for me. Thanks to my brother and husband who keep pushing me so I can finish this work in time.

Many thanks to all the people who have shared my days during my studies and work.

Special thanks to my grandma. I will miss her forever.

Chapter 1

Introduction

1.1 Overview

Face recognition, as a branch in the fields of computer vision, pattern recognition, biometric recognition and neuroscience, refers to verification or identification of a human being based on the visual features of a face.

Face recognition has established its importance via its wide range of applications such as passport verification in customs, identity verification in bank systems, video surveillance in security systems and etc... In Canada, ICBC uses face recognition software[29] to help keeping drivers records in BC province; In Mexico the government adopted *FaceIt*® face recognition technology[20] to eliminate duplicate voter registrations in presidential elections[19, 56]; around the world, face recognition systems are applied in economic entities, entertainment, homes or small appliances for security/entertainment purposes[38, 54].

As face recognition is a practical yet popular research field, researchers have proposed a variety of approaches lending face recognition techniques a high level of maturity. However, we have to admit that the human perception system still remains

mysterious to science research. Consequently the state of the art face recognition algorithms mostly follow mathematical methods other than simulating the biological function of human brains. From the algorithms perspective, face recognition task is usually performed by the comparison of two face images to determine whether they belong to the same individual. Before comparison, a fundamental step for most algorithms is aligning the two face images.

Alignment is known to be a key factor to the face recognition algorithm for its considerable influence on the recognition rate. It is, however, also a thorny issue for which researchers have never come up with a precise definition[53, 62, 52, 51].

Affected by a variety of factors such as facial expressions, facial makeup, pose angle, image quality etc., neither is it possible for a perfect pixel-to-pixel alignment between two images, nor is this perfect alignment ideal or necessary for research needs.

Admitting this, we revise the definition of ideal alignment to that a good alignment does not focus on the best overlap of two images such that the rightmost corner of the mouth in one image is exactly the same position in a coordinate system as it is in the other image, but that the alignment best describes the features of the object to recognize: it should be tolerant of the deviation among images from the same person and tell the difference between images from different persons. The alignment task under this definition is now an alignment that approaches the ideal alignment as close as possible.

An immediate benefit of a better alignment is a relatively accurate description of the offset between two images, and also a higher recognition rate of the face recognition system. The pursuing of a better alignment contributes to a high performance recognition algorithm and thus becomes an important motivation for research, including this work.

1.2 Research Objective

The objectives of this research work are to:

1. Design an alignment scheme that finds a relatively better alignment of two face images;

2. It should be generally adoptable prior to many face recognition approaches to improve their performances;

3. Apply voting theory to the face recognition field and test the performance of hard combination and soft combination for face recognition;

4. Develop an executable framework that integrates our approach and the existing face recognition approaches to evaluate our approach;

5. Take further experiments to test its extensibility.

1.3 Contributions

This work presents an innovative displacement-based aligning scheme that has a high portability to various descriptors and significant improvements in their performances. The greatest contribution of this alignment scheme is that it takes into consideration the regional deviations and dynamically simulates a relatively better alignment for a particular pair of face images.

Regional voting theory is adapted to face recognition problems and has proved its strength for system stablility against image deviations/offsets/noises.

The block LBP displacement-based local matching approach reports outstanding experimental performances in comparison with the original LBP approach.

Experiments demonstrate that our approach applies to not only large-sized images but also small-sized images, and that it also applies to descriptor approaches other than LBP.

Part of the contents of this thesis has been published in [17, 10].

1.4 Thesis Objective

The aim of this thesis is to provide a full view of our algorithm. The following objectives are realized to attain our goal:

1. Introduce the face recognition problem: investigate the face recognition system and discuss the factors that influence the performance of a face recognition system;
2. Review the state-of-the-art achievements in related fields, including those that inspire our approach;
3. Propose our approach;
4. Implement the research design; perform experiments and report experimental outcomes;
5. Refine the algorithm;
6. Test extensibility and report experimental outcomes.

This thesis is an expansion of these objectives and is organized as following:

Chapter 2 explores the face recognition problem and the processes of a face recognition system.

Chapter 3 provides a literature survey of some popular face recognition approaches and voting scheme studies. In particular Chapter 3 gives a full description

for the LBP approach, which we choose as our representative descriptor to perform experiments; and a full description of the study on regional voting, which contributes one of the most important inspirations of this work. Readers with related background can skip Chapters 2 and 3.

Chapter 4 presents our proposed approach.

Chapter 5 shows the experimental results and comparisons with some popular approaches for performance evaluation.

Chapter 6 discusses the extensibility of our approach to descriptors other than LBP, followed by a conclusion in Chapter 7.

Chapter 2

The Face Recognition Problem

A face recognition task is to *verify* or *identify* a person by the facial features.

Depending on the task objective, most face recognition problems fall into two categories: verification and identification. The former is a one-to-one problem: given a face and an identity, determine whether this face comply with the claimed identity while the latter is a one-to-many problem: given a face, the system needs claim its identity from known identities or claim that the identity is unknown.

From a general point of view, specific tasks and applications have been extensively studied in face recognition, such as facial expression recognition, gender recognition, skin texture recognition etc.. The source images can be from 2D images, 3D models, videos, software developed pictures or other sources.

Our study focuses on recognition of 2D images. The following discussion is within our study focus.

A *face recognition system* is a system that performs the face recognition task. It is usually constructed of three components: a gallery set, a probe set, and the recognition component. A *gallery set* is a set of *gallery images* with recognized identities registered with the system. To the understanding of the system, the

gallery image(s)[1] is the only knowledge and the standard description of its associated identity. A *probe set* is a set of *probe images* to be identified or verified. Sometimes a system does not store the probe set, instead, it intakes the probe image at a face recognition request. The *recognition component*, in a face verification task, takes in a probe image and its claimed identity, retrieves the gallery image(s) of the claimed identity, and compares this probe image with the gallery image(s) to make a positive or negative decision. In a face identification task, the recognition component takes in a probe image and compares it with every gallery image to determine the probe image's identity or claim it does not recognize this probe image.

A face recognition system that performs the above mentioned tasks usually follows 4 steps:

Step 1: Image capture

Step 2: Face detection

Step 3: Image normalization

Step 4: Face Recognition

Figure 2.1 shows the processes to prepare an image for face recognition.

(a) image capture (b) face detection (c) normalization

Figure 2.1: Image processed for face recognition

[1]There may be more than one images for the same identity in a gallery set.

2.1 Image Capture

We assume that a face recognition system has its gallery set already stored in its memory, either pre-taken or exported from an existing database; the probe image, on the contrary, is usually taken at the recognition request. The image is then fed to the system to perform the recognition task. Recognition is to perform operations on the images. Before we go further into the next step, we need explore several characteristics of images that affect the performace of a face recognition system.

2.1.1 Digital image

Images, as a manifestation of data, usually come in two forms: analog images and digital images. An *analog image* is continuous in tones with progressive changes, such as a photograph developed from the film or paint on canvas. A *digital image* is discrete, numerical representation stored as a matrix in a digital storage like a portable disk. An image in the computer storage system is always a digital image. A digital image can be taken by a digital camera, scanned from a photograph, projected from a 3D image, captured from a video or created by a graphical program. Analog images can be digitized by technical methods such like scanning.

The two categories of digital images are *vector images* and *raster images.* The former are mostly created by a graphical program based on vectors/functions while the latter are usually based on the dots, the smallest component that constructs an image, which is called a *pixel.* In a face recognition system, the probe image usually is a raster image. A raster digital image is characterised by following features.

An image can be one of the three *color modes*: binary, greyscale or color. In a particular color mode, a number of bits are used to represent the tones of each pixel. This number is called the *bit depth* or *pixel depth.* A bit depth of n yields 2^n

tones. For example, a binary image needs one digit, 0 or 1, representing two colors, typically black and white; a greyscale image usually takes 2 to 8 bits. In a color image, 24-bit is called true color and 30-bit or higher is called deep color, both of which are sufficient to represent images as real for human eyes. The greater the bit depth is, the more tunes the image can represent. The binary value of the bits defines the *pixel value*. An image is represented as a matrix of pixel values.

In a color image, pixel value is understood by the computer under certain *color model*. Most color models are either subtractive or additive mixing. Some famous color models are RGB, CMY, CMYK, HSV and HSL.

A big concern about a digital image is the storage it requires, which closely relates to its *image size*: the number of bits it takes to represent the image. Image size is the product of bit depth and number of pixels. The number of pixels is represented by *pixel dimension*, which is the product of number of pixels per column and number of pixels per row. For example, an image containing m pixels per column and n pixels per row is of size $m \times n$ (pixels). Usually the bigger the image is, the more information it describes: high bit depth will give a richer tone scale and high pixel dimension will contain a greater scope or detailed texture.

A digital image is *compressed* to reduce space cost. Depending on the compression methods, number of colors and etc., images can be of various *file formats*, like TIFF, PNG, GIF, JPG, RAW, BMP, PSD to name a few. Preference of file formats varies by task and usually file formats are mutually transformable.

Figure 2.2 shows a 150×130 sized[2] greyscale digital image. For illustration purposes, this image is resized to 20×16 and its pixel value matrix is shown in Figure 2.3. It is stored in a face recognition system as a *·png* file of bit depth 8 and the accompanying pixel values range between 0 and 255.

[2] 150×130 means that this image has 150 pixels per column and 130 pixels per row.

A face recognition system adopts the raster digital image. The recognition component in fact performs pair comparison(s) between two matrices of pixel values.

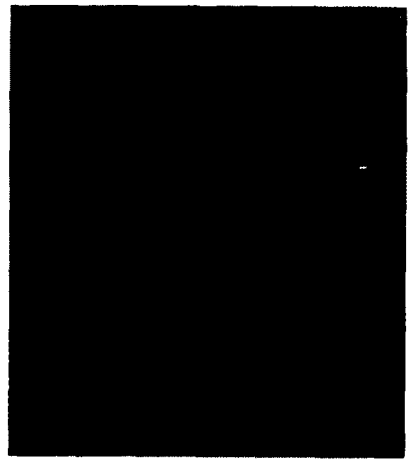


Figure 2.2: A face image

122	38	41	38	40	36	35	53	76	92	109	119	93	59	54	90
79	34	39	39	36	34	61	102	124	125	129	130	113	66	43	52
40	36	37	37	32	58	95	128	135	134	135	131	121	83	40	40
37	35	33	33	51	92	116	133	134	137	144	145	134	104	46	37
37	31	30	38	75	106	117	122	121	121	118	108	117	113	55	33
34	28	35	74	95	95	100	112	108	98	98	115	129	121	68	30
31	30	56	85	81	95	99	112	108	99	97	83	107	126	90	33
29	44	78	82	75	89	101	109	115	98	86	77	99	121	100	31
29	37	89	119	120	122	109	117	121	111	118	119	132	134	97	29
26	44	103	129	131	121	117	136	141	123	124	129	131	136	104	27
24	46	114	129	125	107	113	138	144	120	112	128	131	140	100	24
22	44	116	125	114	102	108	108	108	113	105	117	123	129	89	22
21	37	113	115	110	113	101	90	86	108	117	106	114	120	75	20
22	25	100	119	110	91	98	90	95	103	87	96	121	118	55	21
24	17	67	122	111	81	96	112	110	87	87	111	118	99	31	23
25	19	26	91	116	112	93	93	97	96	115	118	108	54	21	24
27	22	19	34	100	121	108	91	92	116	125	112	71	20	25	24
28	23	21	21	68	109	121	115	117	124	112	85	51	22	25	25
29	25	21	21	64	91	99	106	107	99	85	80	53	23	24	26
29	29	24	22	54	97	95	92	91	89	86	85	51	23	24	28

Figure 2.3: 20×16 sized pixel matrix of the image in Figure 2.2

2.1.2 Taking a photo

As we may easily claim that the more information the image contains, the higher recognition rate a system achieves. However, it is not true. We can simply learn this by thinking over how many times we took a high resolution picture that did not look like ourselves at all. Under this observation, one question arises immediately is: How can we take a picture that best describes our face? This may be answered differently from the aesthetic point of view or with concerns of face recognition rate. We here discuss several key influential factors that could be helpful to improve the performance of a face recognition system.

Photographic equipment is the first choice we make in research studies because we need to provide the parameters of the equipment we use to collect the data. Then follows *image size*. It seems that big-sized images (or big-sized faces to be precise) should always be preferred since it offers more information than smaller ones. However, bigger image size also requires long processing time due to its large number of bits. In an image processing system, such as a face recognition system, the trade-off between the image size and processing time is taken as the trade-off between accuracy and efficiency. Some images, like the medical images, require highly detailed information while others might call for a faster processing time. Such a trade-off should take under consideration the emphasized system features and task requirements.

Pose angle is a big concern for face recognition. The best angle of a picture taken for face recognition is the frontal image as it covers the wholė region of the face. Pictures taken with an angle, horizontally, vertically or an arbitrary angle may cause absence of data while it is believed that full information of both sides of the face is helpful for the recognition decision as most of the human faces are not strictly symmetric. Also that pose angle causing different facial regions fall

into different focal length to the lens will frequently result in distortion of the face[3] and illumination changes is a most frequent accompanying sideffect of post angle. Research shows that the recognition rate decreases as the pose angle increases, especially when the horizontal angle is greater than 30 degrees or the vertical angle is greater than 15 degrees [24].

Illumination, as studied in many research works, can greatly lower the recognition rate[41, 1]. That is to say, the change induced by illumination could be larger than the difference between individuals. Some face recognition approaches perform stably against illumination change, such as LDA. Some approaches apply strategies, such like histogram equalization, to reduce the illumination effects. We believe an illumination-oriented method should not be the final solution for a face recognition system, given that in the real world the image distortions vary and most likely are a result of a combination of many factors.

(a) pose angle (b) illumination

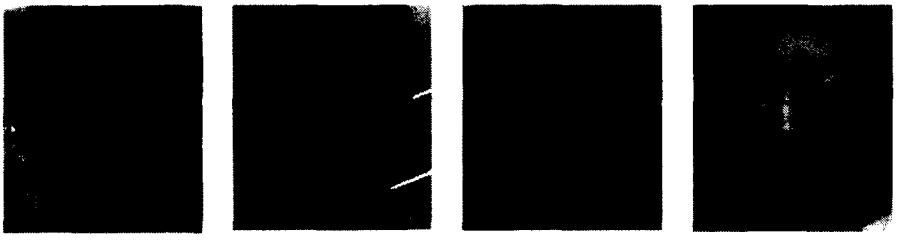
Figure 2.4: Images taken with different angles and illumination conditions

A possible solution against pose angle and illumination is that by a fine control of shooting conditions, we may strictly restrict the influence to a small scale to take most comparable images of a person; however, this fails to deal with uncontrollable circumstance such like a video surveillance image taken under any illumination from

[3]Cosmetic guides suggest a 45-degree angle depression to give a look of skinner cheeks and bigger eyes to make a doll-like face.

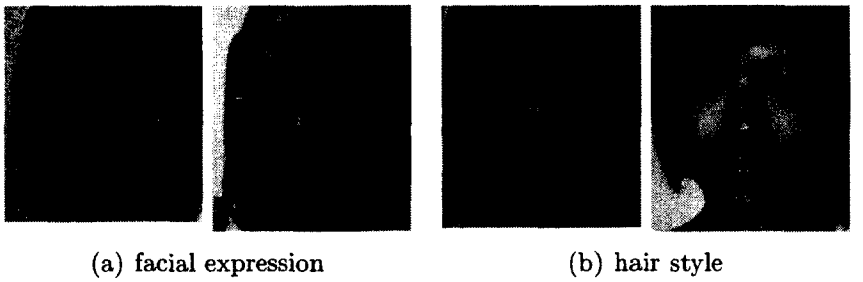
any angle, or images that are from different sources where it is infeasible to unify everything, such as photos taken at airports around the country.

There are also irresistible changes happen to human faces such as facial expressions, aging, pimples and scars, makeups, apparels, cosmetic surgery, hair styles(hair growth), glasses, rings, color contacts and others accessories that people wear... A mature face recognition system should not refuse these changes as they are taken as part of the features of a human face.



(a) 2005 (b) 2006 (c) 2010 (d) 2012

Figure 2.5: Images taken at different time



(a) facial expression (b) hair style

(c) glasses (d) makeup

Figure 2.6: Images deviations

If an image is scanned from a photograph, machinery may cause *image noise* from an unclean surface or *distortion* from a warped or wrinkled original file.

From above, the performance of a face recognition system relies heavily on the images. A face recognition system in practice may focus on specific factors for particular tasks while the study of the face recognition approaches should take into considerations all possible factors. A good recognition approach should minimize the intra-class difference and maximize the extra-class difference, that is, be stable against the biological feature distortion of human faces and image noises while remaining sensitive to the difference between different individuals. We further expect it be reasonably stable against a non-standard image with an angle, an unfavorable illumination or different image sizes.

2.2 Face Detection

Face detection, as a pre-process for the face recognition task, is itself a research field in object-class detection. It aims at finding faces in an image taken under any condition where there can be none, one or more faces. Sometimes the faces are processed by rotating, scaling or other means if the face in the image is not in a preferred position[4].

In face recognition task, face detection is to find the location and size of the face and excludes background (non-face areas) from the image. Free face detection software include Facial landmark detector(Center of Machine Perception, Czech Technical University, Prague), face detection using support vector machine(SVM) (Omid Sakhi), FDLIB (W. Kienzle and etc.); companies like ACSYS, Betaface, Luxand offer commercial ones too[22].

[4]One example is when a boy does handstand, his face is upside down.

2.3 Face Normalization

Face normalization prepares the images for comparison. It comes in two forms: geometric normalization and image condition normalization.

Geometric normalization asks for a unified image size with fixed facial feature positions and scales (such as fixed centers of the eyes, distance between eyes[5], the middle of the upper lip and other believed key facial features). This is achieved by clipping, resizing, scaling or rotation if necessary. Image condition normalization pre-processes images with unfavourable parameters such as lighting or contrast. This can be done by global filtering, local modification, histogram modification or with a lighting compensation mask.

Advanced features of face normalization include facial expression normalization, facial orientation and many others by global or local modifications. Technical means are adopted to minimize the deviations caused by image conditions.

As face detection and face normalization can be performed separately from face recognition, some recognition systems do not take this task into consideration but solely focuses on recognition. According to their different attitudes towards the two procedures, face recognition systems fall in two categories: the ones include face detection and normalization are called fully automated systems and those do not include the two procedures are called partially automated systems or semi-automated systems.

[5]Figure 2.1(c) is normalized with distance between centers of eyes being 56 pixels and the centers of eyes lies on the 53th pixel of the same column.

2.4 Face Recognition

Normalized images are ready for comparison. The result of comparison is the answer to the identification or verification task. In the literature, many approaches are developed for face recognition purposes with various and reasonable emphases, namely feature-based methods, appearance-based methods, descriptor-based methods, template-based methods, and neural network methods. For example, in a feature-based method, facial features are extracted from the normalized faces to get the nose, eyes and other believed to be important features, and then the feature vectors from two images are compared to derive a final conclusion. In a descriptor-based method, a descriptor is applied to get a description of the face, usually in the form of a vector. Then comparison takes place between vectors. A detailed discussion of the comparison methods will be presented in the Literature Review.

Chapter 3

Literature Survey

Face recognition is easily seen as a bionics application as it is never difficult for a human being to recognize an acquaintance; however, it remains unknown how our brain performs such a task. Biologists and engineers keep exploring and have made many insightful yet interesting observations.

Wilmer et al[55] found that human face recognition ability is specific and highly heritable by observing "correlation of scores between monozygotic twins (0.70) was more than double the dizygotic twin correlation (0.29)" and that "low correlations between face recognition scores and visual and verbal recognition scores indicate that both face recognition ability itself and its genetic basis are largely attributable to face-specific mechanisms"[55]. Similar observations have been made in many studies supporting that the brains have a specific section to perform the face recognition. A good evidence for this claim might be the face blindness disorder (prosopagnosia)[23], in the study of which the fusiform face area[31] is believe to be specialized for face recognition. A model built by Haxby et al further suggested that facial identity and expression might be processed by separate systems[25, 42].

Young et al[60] drew their conclusion that facial features were processed holisti-

cally from an experiment in which subjects found it more difficult to recognize the faces when the top half of one face is combined with the bottom half from another face[60]. Sadr et al made the interesting observations that recognition performance for faces without eyebrows was significantly worse than that for faces without eyes, which listed the eyebrow a key facial feature for face recognition that is no less important than the eyes[36].

Though extensive studies have been carried out to satisfy the curiosity for human face perception system, there is still a long way to go before we could translate the human perception manner into a computer-based algorithm and lead artificial face recognition in a biotic manner. Present research, other than simulating the biological functions of human brain, remains algorithmic in manner.

Artificial face recognition algorithms base recognition on comparison of paired face images. Varied in how the images are processed, most traditional algorithms adopt one of the two approaches: holistic approaches and regional approaches. The former calculate the image as an integrated input while the latter breaks the image into regions. In recent years a new scheme arises by adopting the voting theory to face recognition, referred as the regional voting approaches.

The research on face recognition initiates with holistic approaches in the late 1980s. Holistic approaches take an entire human face as a numeric matrix which is converted into a vector in multidimensional space by concatenating the rows of matrix one after another. These face vectors are then projected into lower dimension spaces for similarity measurement. Different approaches vary in their methods of projection (standard projection, differential projection or kernel Eigenspace projection). Examples of holistic approaches are the Eigenspace-based approaches such as Principle Component Analysis (PCA)[43][49], a later 2D-PCA[59], Fisher Linear Discriminant (FLD)[6], Evolutionary Pursuit (EP), Linear Discriminant Analysis

(LDA)[21], Independent Component Analysis (ICA) and etc.[2, 37, 50, 58, 5].

In the mid-1990s, research tends to focus on the different contributions of different regions from the face and thus led to a blossomming in the study of regional approaches. Regional approaches break the face into regions, aiming at preserving locality from which more discriminating face features would be used for comparison. Examples of regional approaches include subpattern PCA (SpPCA), Elastic Bunch Graph Matching (EBGM), Local Binary Pattern (LBP), Local Gabor Binary Pattern(LGBP), and Histogram Sequence (LGBPHS)[18, 3, 4, 47, 63].

For regional approaches, one thing worth mentioning is that when an approach extracts discriminative information locally from the face, shall it emphasize the biological features of the face, resulting the regions representing the facial features, or shall it emphasize the layout of the face features, resulting the regions representing portions of the face. Based on the assumption that some region might have more influence to identify a person, weight-scheme can be put to the regions to represent this property. Weights can be assigned based on the educated guesses such as that eyes and eyebrows are more discriminating than the cheek or forehead; or they can be empirical values that come out of the training process, if any. Regardless of whether focusing on features or spatial layout of the features, an accompanying concept that comes with many regional approaches is the descriptor, with which, a standard input face image is processed to a representation generated by this descriptor to better serve the calculation.

A new category of approaches that arose lately is regional voting approaches, which is more a general scheme[11, 12, 13, 14] that could apply to many research fields other than face recognition[15, 16, 9]. The main objective of introducing the regional voting scheme is to create a system that is more stable against noise[14]. The voting theory applies to face recognition in such a way that voting scheme

prevents the system from changing its decision based on the facial changes caused by aging, illumination changes or other irresistible influences.

3.1 The LBP Approach and Its Variants

3.1.1 LBP approach

Local Binary Pattern(LBP) is a regional descriptor-based approach originally proposed by Ojala et al for texture description[45, 44] and later introduced to face recognition.

LBP works as follows:

Given a face image, an LBP operator applies to obtain its LBP map by thresholding P sampling points on a circular neighbourhood of radius R centered at a pixel. Depending on the value of the center pixel and that of its neighborhood, a binary number 0 or 1 is assigned to its neighborhood representing whether the pixel value of this neighborhood is less than or greater than or equal to the center pixel value. Then the concatenation of the P binary values is taken as the label of this center pixel and all labels construct the LBP map of this image.

The LBP map is then divided into windows. In each window, a histogram representing the distribution of the numerical labels for pixels in this region is generated to be the texture descriptor of this region and histograms from all windows are concatenated to form the LBP description of the whole face image.

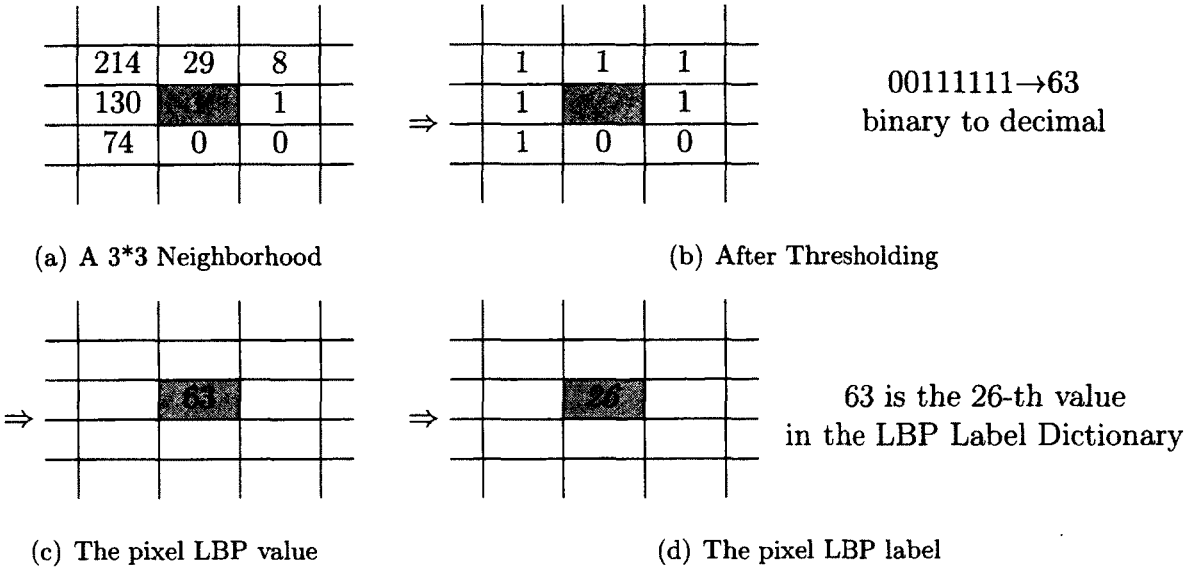
Figure 3.1 shows a basic LBP operator. Figure 3.1(a) is a 3×3 area from Figure 2.2. To calculate the LBP label for the pixel in the center, by thresholding 8 sampling points on a circle of radius 1, we obtain an eight-bit string, which, if counted anticlockwise from the bottom right one, equals 63 in decimal as in Figure 3.1(c).

A uniform pattern in LBP is defined as an eight-bit string which contains at most two bitwise 0/1 transitions if examined circularly, i.e. it is a circular concatenation of a series of 0s and a series of 1s. An eight-bit string has 57 uniform patterns. The LBP label dictionary is a vector of 58 elements, containing 57 uniform patterns and 1 non-uniform element, as shown in Figure 3.2. Under this definition, the LBP label in Figure 3.1(c) will be labeled as in Figure 3.1(d).

Ojala et al[44] observed in their experiments that uniform pattern in texture images counts for about 90% and 70% for using 8 sampling points on a neighbourhood of radius 1 and 16 sampling points on a neighbourhood of radius 2 respectively and proposed to classify only the LBP labels in the uniform patterns and classify all non-uniform ones into one category.

In a more complicated form, LBP operator can have different radius with different sampling points that evenly distributed along the circular neighborhood. An LBP operator with P sampling points of radius R is denoted as $LBP_{P,R}$. When a sampling point does not fall into the center of a pixel, bilinear interpolation is adopted to find the value of the sampling point. The LBP operator with consideration of uniform patterns is denoted as $LBP^{u2}_{P,R}$.

LBP has reported high performance by maintaining three levels of localities: the labels on a pixel level, the histogram representation on a regional level and the concatenated histograms on a global level. As we believe regional approaches should outperform many holistic approaches, and LBP is one of the reported best performing regional approaches, we come to the choice of applying our scheme to this LBP approach.



(a) A 3*3 Neighborhood

(b) After Thresholding

(c) The pixel LBP value

(d) The pixel LBP label

Figure 3.1: A basic LBP operator

String	Decimal	LBP label
00000000	0	0
00000001	1	1
00000010	2	2
00000011	3	3
00000100	4	4
00000110	6	5
00000111	7	6
00001000	8	6
00001100	12	7
00001110	14	8
00001111	15	9
00010000	16	10
00011000	24	11
	⋮	
11111110	254	56
11111111	255	57
non uniform		58

Figure 3.2: LBP dictionary

3.1.2 TPLBP

Three-Patch LBP (TPLBP) was introduced by Wolf et al[57] as a variant of LBP. It works as follows:

A patch C is defined as a $w \times w$ region centering on a pixel c. TPLBP, as is named, involves three patches to calculate a bit code which later contributes to the TPLBP code for this pixel.

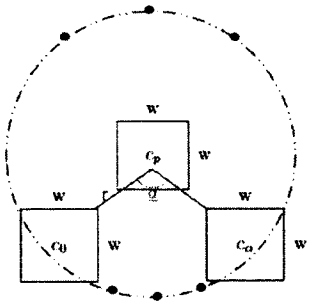
For any pixel c_p, TPLBP first finds the patch C_p and S patches $C_i (i \in \{1, 2, \cdots, S\})$ distributed evenly along a circle of radius r from c_p. A pair of patches are two patches that are α patches apart along this circle. One bit code is generated by thresholding the difference of distances between C_p and each patch from the pair, and the TPLBP code for the center pixel is a concatenation of all bit codes.

A TPLBP operator in a general form is denoted as $\text{TPLBP}_{r,S,w,\alpha}$ as shown in Figure 3.3. Defining a thresholding function $f(x)$ in Equation 3.1, the TPLBP code for p is given in Equation 3.2.

$$f(x) = \begin{cases} 1, & x \geq \tau \\ 0, & x < \tau \end{cases} \tag{3.1}$$

$$\text{TPLBP}_{r,S,w,\alpha}(p) = \sum_{i=0}^{S-1} f(d(C_i, C_p) - d(C_{(i+\alpha) \bmod S}, C_p))2^i \tag{3.2}$$

An image represented by its TPLBP codes is then divided into non-overlapping regions and a histogram representing the distributions of the codes is first generated for this region and then normalized to unit length. The concatenation of the normalized histograms constructs the TPLBP representation of this image.



the 0^{th} bit code for c_p generated by this $\text{TPLBP}_{r,S,w,\alpha}$ is:
f(d(C_p,C_0)-d(C_p,C_α))

Figure 3.3: A TPLBP operator

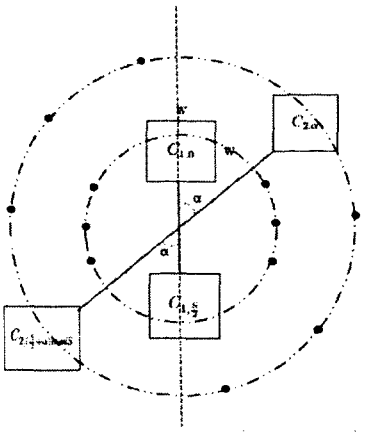
3.1.3 FPLBP

Followed TPLBP, Wolf et al proposed another variant of LBP, named Four-Patch LBP (FPLBP)[57]. Similar to TPLBP, FPLBP bases the bit code of c_p on four patches distributed evenly along two circles of radii r_1 and r_2, with two center symmetric patches on each circle and the inner patches apart α patches from their corresponding outer patches respectively.

Let $C_{1,i}$ denote the i-th patch on the inner circle and $C_{2,i}$ denotes the i-th patch on the out circle. With the same thresholding function $f(s)$ given in Equation 3.1, the $\text{FPLBP}_{r_1,r_2,S,w,\alpha}$ is given in Equation 3.3.

$$\text{FPLBP}_{r_1,r_2,S,w,\alpha}(p) = \sum_{i=0}^{S/2} f(d(C_{1,i}, C_{2,(i+\alpha) \bmod S}) - d(C_{1,(i+S/2) \bmod S}, C_{2,(i+S/2+\alpha) \bmod S}))2^i \quad (3.3)$$

The global TPLBP description of the image is generated following the same process as in TPLBP.



the 0^{th} bit code for c_p generated by this $\text{FPLBP}_{r_1,r_2,S,w,\alpha}$ is:
$\text{f}(\text{d}(C_{1,0},C_{2,\alpha})\text{-d}(C_{1,S/2},C_{2,(S/2+\alpha) \bmod S}))$

Figure 3.4: An FPLBP operator

3.2 Regional Voting

The stability of regional voting was proved by Chen and Tokuda in 2003[12], and later introduced to the studies of face recognition. To learn this voting scheme, we will first have a glance at the voting problem in general.

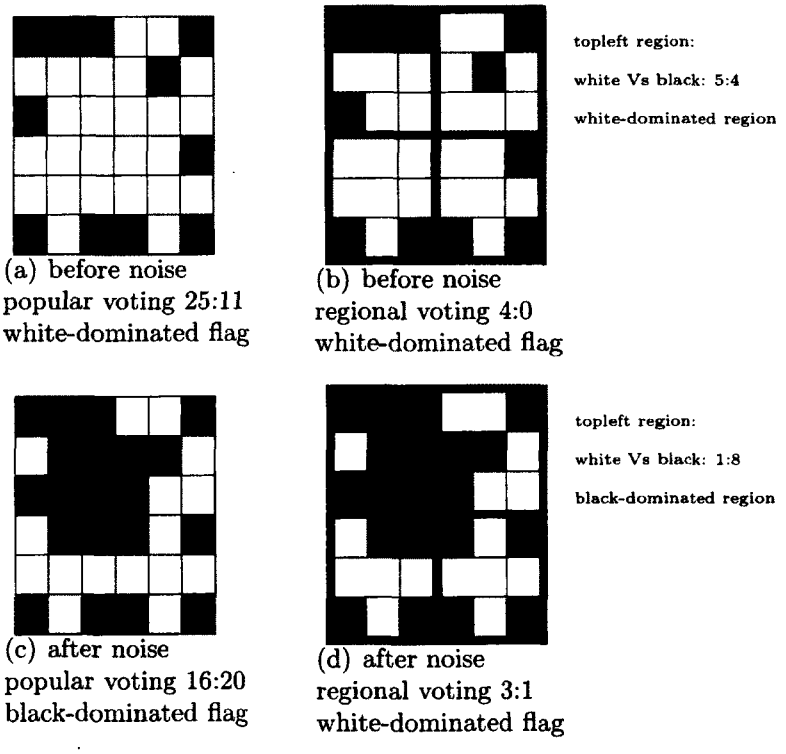
A voting problem is to ask a population of $\mathcal{M}$ to select one winner out of $\mathcal{N}$ candidates. This selection can perform in two manners: direct popular voting and regional voting.

In direct popular voting, the $\mathcal{M}$ voters each draws a vote to one of the $\mathcal{N}$ candidates and the candidate who gets the most votes wins. In regional voting (also called local voting or Electoral college), the $\mathcal{M}$ electors are first grouped into $\mathcal{X}$ regions, and a direct popular voting for the $\mathcal{N}$ candidates takes place within one region to generate a local winner on a winner-take-all basis. Later the regions, each performing as a single voter, vote for the final decision and the candidate who gets the most votes from the $\mathcal{X}$ regions wins.

Noise is introduced to study the stability of voting schemes. A noise refers to

a sudden change of the decision from one vote and the stability is watched against noise. A system is said to be stable if the final result remains against noise.

A simplest form of the voting model developed by Chen and Tokuda is illustrated in Figure 3.5. A binary flag of size 6×6 is used to represent the 36 voters in a two candidate voting; a white pixel means this voter votes for candidate A and a black pixel means this voter votes for candidate B. Figure 3.5(a) and Figure 3.5(c) show a block of noise turns over the original decision in a direct popular voting and Figure 3.5(d) demonstrate a regional voting retains the original decision confronted with the same noise.



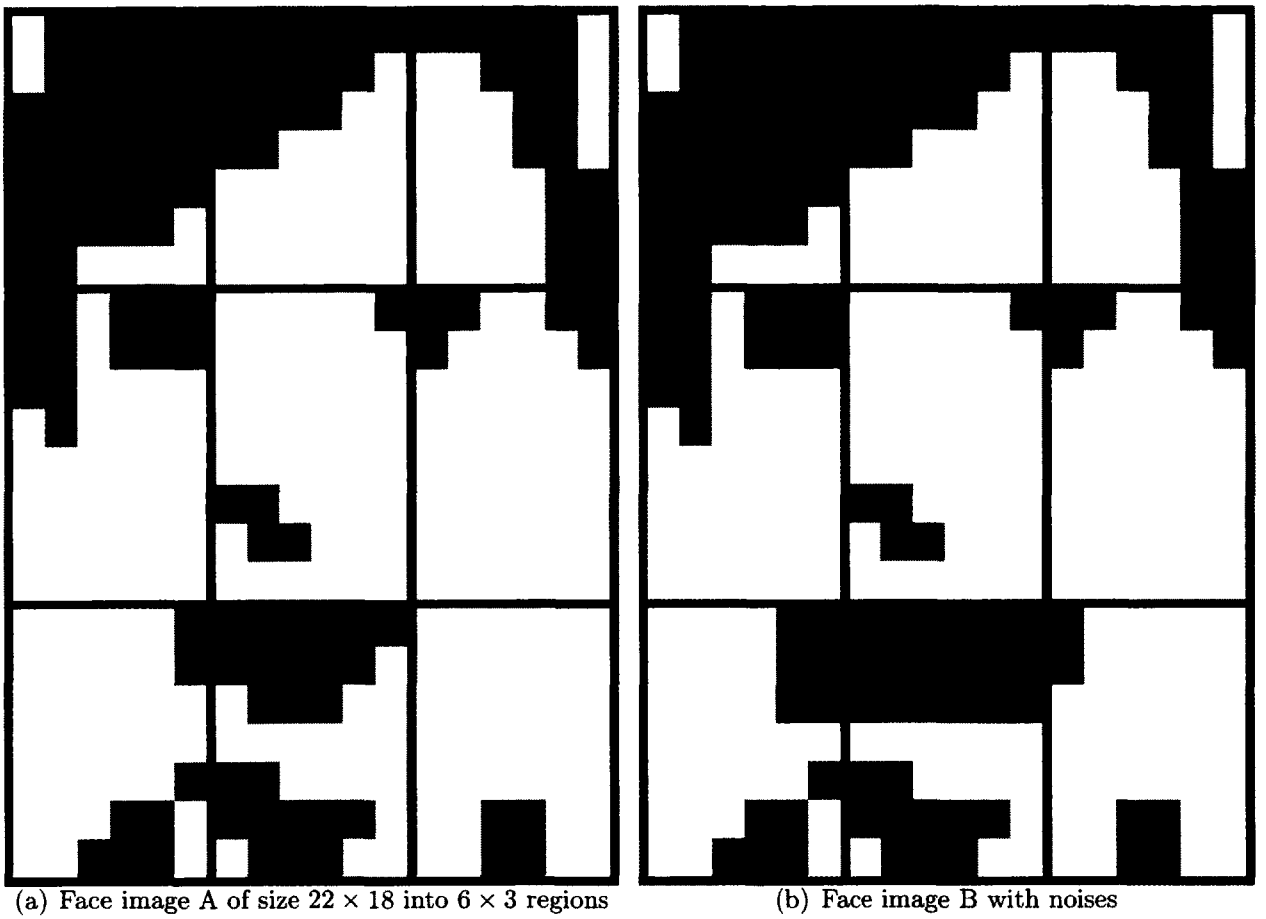
(a) before noise
popular voting 25:11
white-dominated flag

(b) before noise
regional voting 4:0
white-dominated flag

(c) after noise
popular voting 16:20
black-dominated flag

(d) after noise
regional voting 3:1
white-dominated flag

Figure 3.5: The flag model for voting



(a) Face image A of size 22×18 into 6×3 regions

(b) Face image B with noises

Figure 3.6: Regional voting in face recognition

Chapter 4

Proposed Algorithms

A deep look into the literature gives us the understanding that holistic approaches and regional approaches both take the whole face image as the input and each pixel, regardless of whether it is represented by its pixel value or some other value generated by a descriptor, contributes evenly likely to the final decision (even in a weighted scheme, the pixels in same-weighted regions contributes the same). We believe, as supported by [12, 14], such approaches *lack tolerance for what is called noise* (the sudden change of the values of some pixels which makes these pixels no longer corresponding to its original objective), *neither do they enhance the ability to tolerant the biological deviations of face features* which might happen only in some random regions of the face area and expose the disadvantage of fixed-weight scheme. In fact, a pre-set alignment always draws defects in some cases. Regional approaches work better, but still both holistic and regional approaches fail to deal with this issue.

These weaknesses of pre-set alignment schemes and fixed weight schemes lead us to the conception of a scheme that could dynamically locate a best alignment by simulating all possible alignments corresponding to all possible deviations of facial

regions to conquer the alignment issue. Our scheme is established on the following observations.

First, admitting the existence of deviations means that a deviation is consistent with the real-world face features and different regions may (usually they do) have different deviations: sometime the forehead in the probe image has a positive deviation from the forehead in the gallery image while the mouth in the same probe image has a negative deviation from that in the gallery image. Thus the common regional methods of assuming that all regions have the same deviation may seem lose precision. Simply cutting an image into regions and combining them back into a face gains no advantage besides locality of facial features. We believe different deviations should exist for every pair of corresponding regions from the galley image and the probe image. Deviation between each pair should be found dynamically rather than assuming they are aligned to their original location in the image by default. This dynamic aligning process is much like shifting one image/region around the other to find a better position to align the two. We therefore propose a framework that simulates every possible deviation in units of a pair of corresponding regions to conquer this problem.

Second, even a great change within a small region (corresponding to concentrated noise[12]) should not overturn the final recognition decision under circumstance where most of the other regions remain the same. It also means that even when there is a light change spread over every region (corresponding to salt-and-pepper noise[12]), as long as the regions retain this person's identity, the final recognition decision should remain unchanged. For example, one might have temporary blood scabs on the chin and forehead from a car accident, which results in the similarity related to these regions to be extremely low, even further denies the final decision; while we intuitively perceive that the scabs should only be reflected on the decision

of their region(s), leaving decisions of other regions unaffected. However, even in regional methods, changes caused by facial features deviation or image noise within one region will cause a change in the description of the whole face and thus result in a different similarity between two images. We believe we could gain system robustness by constraining regional deviation and regional noise within its own region by applying voting theory to our scheme.

Regional voting scheme is the right solution that meets our goal: system robustness against noise. A face image fits in the regional voting scheme easily if each pixel is taken as a voter and the identities are taken as candidates. A face verification task is a one candidate voting where pixels vote for "positive" or "negative" to this candidate and the final decision is positive if the majority of pixels vote for positive or otherwise; A face identification task is a multi-candidate voting where each registered identity is a candidate and the pixel votes for one candidate and the final decision is the identity that gains the most votes. In two images, a noise happens if the pixel value in one image does not equal to its corresponding pixel in the other. In the theory part, noise corresponds to deviations or facial features changes. As an example, Figure 3.6 on page 28 shows two images of the same person and Figure 3.6(b) shows a noise contaminated image caused by smiling.

The regional matching approach is adapted to face recognition in such a manner that a face image is divided into blocks (representing one region in the face) and decision from each block is made on statistics within this block which later votes for the final recognition decision. We then can take benefit from the voting scheme to construct an algorithm that is more stable against the deviation of face regions. Another immediate benefit is its stableness against noise. There are many reasons causing noise in images, such as regional shading from not preferred photographing or an unclean scanner surface when digitizing a filmed photo. As [14] suggests, re-

gional voting gains robustness against both concentrated noise and salt-and-pepper noise.

As our intention originates from building an aligning scheme that may fit more than one comparison methods, the literature survey leads our attention to the descriptor based approaches, then the LBP descriptor to be particular[64, 30, 3]. LBP descriptor outperforms many state-of-the-art face recognition approaches by its high descriptive localities[3, 4], and a comparison between the original LBP and the regional matching LBP should be capable of exposing the advantage of regional matching schemes over other descriptors. Also, given that LBP has reported high recognition rate[3], it should be interesting to see whether we can go further and how far we can go in artificial face recognition.

Integrating the above proposed ideas, we come to the conception of constructing a template framework that could dynamically generate all possible alignments from which we locate the best alignment based on displacement in units of image regions. LBP is used as the descriptor and regional voting is adopted to construct a displacement-based local matching approach, we name LBP-DLMA. We expect a high portability of this template to apply to any descriptor based matching approach. Further more, for a comprehensive framework, various descriptors can apply to the regions and thus a higher recognition rate is expected by taking advantage of the different descriptors.

LBP-DLMA works as follows:

Given a face image from the database, we first generate its LBP map. Believing that deviations vary among pairs of corresponding local regions from two images, we partition this LBP map into blocks. By assigning deviation values to blocks enumerately and respectively, we can generate a set of candidate face alignments, which together constructs the template description of the face. A best alignment is

located as one whose similarity is highest among all.

Having located the best aligned face, every block in this best aligned face takes an internal election on a winner-take-all basis to generate the local decision of this block, which contributes one voter to determine the final decision.

4.1 LBP Displacement Concepts

Given a face image, we obtain its LBP map of size $(m+2s) \times (n+2s)$ using the LBP descriptor[1] in [3]. By removing h, i, j, k pixels ($h, i, j, k \geq 0$, $h+i = 2s$, $j+k = 2s$) from top, bottom, leftmost and rightmost margins respectively, we obtain $(2s+1)^2$ slightly smaller LBP maps of size $m \times n$. Each $m \times n$ sized map is called a *layer* and is denoted by $I_i (1 \leq i \leq (2s+1)^2)$.

Then we partition each layer into $K \times L$ *blocks* (K blocks per column, L blocks per row). A block in the r-th row, c-th column of the l-th layer is denoted by $B_{r,c,l}$. The set of corresponding blocks from all layers are called *a pile of LBP displacement blocks*, or *an LBP displacement pile*. The pile of blocks in the r-th row, c-th column is denoted by $P_{r,c} = \{B_{r,c,l} | 1 \leq l \leq (2s+1)^2)\}$.

The set of all LBP displacement piles for a face image generates the *template*, or *the LBP displacement description* of the face image, denoted by $\mathcal{T} = \{P_{r,c,} | 1 \leq r \leq K, 1 \leq c \leq L, \}$. Template for a gallery image is called a *gallery template* and template for a probe image is called a *probe template.*

A *candidate face description*, or *candidate face* for simplicity, is a recombination of the blocks, one from each pile, in the template. The template has $(2s+1)^{2KL}$ candidate faces[2], representing all possible deviations of individual block. Let a *test*

[1] All following mentioned images are LBP maps of the images and to be simple, we use the term *image* referring to *the LBP map of the image.*

[2] Each block is selected from $(2s+1)^2$ blocks of its own pile and a candidate face contains $K \times L$ blocks. Thus the number of candidate faces is $((2s+1)^2)^{K \times L}$, equally $(2s+1)^{2KL}$.

pair be two candidate faces from a gallery template and a probe template respectively. A *best matched pair* can be located by exhaustively testing the similarities of the test pairs and choosing the pair with the highest similarity.

Such a template retains the three levels of localities that the original LBP operator has: LBP label on the pixel level, histogram on the regional level and concatenated histograms on the global level. It also represents three levels of deviations: fixed deviation on a window level, dynamic deviation on a regional level and multi-deviation on a global level. It gains tolerance on the deviations of images from the same person from the three levels of deviations.

As an illustration of this template framework, assuming that Figure 4.1(a) is a 18×14 sized LBP map of the face image in Figure 2.2. Let $s = 1$, by removing 2 pixels from bottom, one pixel each from leftmost and rightmost margins, we obtain a 16×12 sized layer which we partition into 12 blocks, each of size 4×4, as shown in Figure 4.1(b). With different margins taken off, there are a total of 9 such 16×12 sized layers, each of which can be partitioned into 12 blocks. Figures 4.2(a) - 4.2(i) show an LBP displacement block pile consisting blocks corresponding to the shaded block in Figure 4.1(b) from all layers. Note that the second block(Figure 4.2(b)) in the pile is the shaded block in Figure 4.1(b). We have 12 such LBP displacement piles, as shown in Figure 4.3. The union of the l-th ($l = 1, 2, \cdots, 9$) blocks from all piles is the l^{th} layer, a 16×12 sized LBP map obtained by removing i, $2 - i$, j and $2 - j$ pixels from top, bottom, leftmost and rightmost margins respectively, where $0 \leq i, j \leq 2$. The set of all these 12 LBP displacement piles is the LBP displacement description template of this face image.

A drawback of such a "simulate by enumerate" strategy is the time cost. However, we make following observations to reduce the time complexity while retaining the descriptiveness of the template.

The first observation is that duplicate test pairs[3] exist in cases where the margins cut from gallery image and the probe image are the same. To reduce redundant comparisons we restrict that the margin parameters h, i, j, k in the probe template all equal s, yielding a probe template with only one block in each pile and only one layer in this template. This restriction will reduce comparisons[4] of test pairs from $(2s+1)^{4KL}$ to $(2s+1)^{2KL}$. Assuming the time cost for computing the similarity between the standard descriptions (without using template structures) of two face images is O(T), the total time complexity computing the similarity between two description templates will be $O((2s+1)^{2KL}T)$ if we compare all pairs of candidate faces.

Under this restriction, the search for the best alignment would be the search for a candidate face in the gallery template which is best aligned with the probe template which contains only one face. We can define *a best matched face* being the gallery face from the best matched pair.

The second observation is that a best matched face retains its "best match" property over all regions[5]. That is, the candidate face from the gallery template which contributes the best match with the probe template should be a combination of blocks that are locally best matches of their own piles respectively. This observation suggests that not all candidate faces need be tested and only the locally best aligned regions should be under our consideration. By a divide-and-conquer

[3]Two images with the same margins still vary given different values of the margins, however the offset is too small to affect the final result thus we can take them as "duplicate test pair".

[4]A comparison associates with a test pair. A test pair is selected by choosing one candidate face from the gallery template out of $(2s+1)^{2KL}$ choices, and one candidate face from the probe template out of the same choices, yielding $(2s+1)^{2KL} \times (2s+1)^{2KL}$ choices of test pairs , equally $(2s+1)^{4KL}$. By restricting the probe template containing one candidate face, the choices of test pairs is reduced to $(2s+1)^{2KL} \times 1$, equally $(2s+1)^{2KL}$.

[5]Proof: Assume the best matched face $\mathcal{G}$ contains one block B_1 whose similarity with the corresponding block from the probe template is less than another block B_2 from its own pile. Replacing B_1 by B_2, we then have a face $\mathcal{G}'$ whose similarity with the probe template is higher than $\mathcal{G}$, which is contradictory to that $\mathcal{G}$ is the best matched face.

strategy, we can further reduce the time complexity to[6] $O(KL(2s+1)^2T)$.

4.2 Similarity Metrics

Assuming the global LBP-based representations of two face images are $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \cdots\}$ and $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2, \cdots\}$, the typical metrics for calculating the similarities between two global LBP descriptions are [3]: Euclidean Distance[7], Histogram Intersection, Log-likelihood statistic and χ square statistic:

$$\text{Euclidean Distance:} \quad E(\mathcal{G}, \mathcal{P}) = -\sum_i (\mathcal{G}_i - \mathcal{P}_i)^2 \tag{4.1}$$

$$\text{Histogram Intersection:} \quad H(\mathcal{G}, \mathcal{P}) = \sum_i \min(\mathcal{G}_i, \mathcal{P}_i) \tag{4.2}$$

$$\text{Log-likelihood Statistic:} \quad L(\mathcal{G}, \mathcal{P}) = \sum_i \mathcal{G}_i \log \mathcal{P}_i \tag{4.3}$$

$$\chi \text{ square Statistic:} \quad \chi^2(\mathcal{G}, \mathcal{P}) = -\sum_i \frac{(\mathcal{G}_i - \mathcal{P}_i)^2}{\mathcal{G}_i + \mathcal{P}_i} \tag{4.4}$$

Such metrics can also be used to calculate the similarity between the block level LBP descriptions of two blocks. As [3] suggests that the log-likelihood measure is not appealing for face recognition, we shall not use it as a similarity measure in this work. Note that in each block, there are one or more windows; the block level LBP description for a block is the concatenation of window level LBP statistics.

[6] A best aligned block is found within its own pile out of $(2s+1)^2$ blocks and there are $K \times L$ best aligned blocks to find to construct a best aligned face, so the total number of comparisons is $(2s+1)^2 \times K \times L$, equally $KL(2s+1)^2$.

[7] We will use a squared version of Euclidean Distance for simplicity in calculation.

4.3 An LBP Displacement Template Matching Approach: LBP-DLMA

Given a gallery set $\mathcal{G} = \{\mathcal{G}^1, \mathcal{G}^2, \cdots\}$ of size T and a probe image $\mathcal{P}$, we base regional voting approach on the following vote definitions:

Let $\mathcal{G}P_{r,c}$ denote the block pile in the r-th row, c-th column in the gallery template, $\mathcal{G}B_{r,c,l}$ denote the block in the l-th layer from this pile, and let $\mathcal{P}B_{r,c}$ denote the block in the r-th row, c-th column in the probe template[8]. $\mathcal{P}B_{r,c}$ also denotes the block pile that contains $\mathcal{P}B_{r,c}$. Assuming the similarity between two blocks is defined as $Sim(\mathcal{G}B_{r,c}, \mathcal{P}B_{r,c})$, with a higher value representing a higher similarity. We then define the similarity between two block piles as in Equation 4.5:

$$\text{SimSim}(\mathcal{G}P_{r,c}, \mathcal{P}B_{r,c}) = \max_{1 \le l \le (2s+1)^2} (\text{Sim}\{\mathcal{G}B_{r,c,l}, \mathcal{P}B_{r,c}\}) \tag{4.5}$$

In a face verification task, one probe template $\mathcal{P}$ is compared with one gallery template $\mathcal{G}$. For each block $\mathcal{P}B_{r,c}$, its optimally aligned corresponding block $\mathcal{G}B_{r,c,opt}$ is the one from $\mathcal{G}P_{r,c}$ that has greatest similarity with $\mathcal{P}B_{r,c}$, also the one that votes for the local identity of $\mathcal{P}B_{r,c}$ following Equation 4.6. A final decision is a *confirmation* or *negation* to the claimed identity, whichever that takes more votes from the blocks.

$$\text{vote}(\mathcal{P}B_{r,c}) = \text{THR}(\text{Sim}\{\mathcal{G}P_{r,c}, \mathcal{P}B_{r,c}\}) \tag{4.6}$$

In a face identification task, one probe template $\mathcal{P}$ is compared with every gallery template $\mathcal{G}^t$ in the gallery set. Block $\mathcal{P}B_{r,c}$ is believed to share the identity of $\mathcal{G}P^I_{r,c}$ which has the greatest similarity among all $\mathcal{G}P^t_{r,c}$ where $t \in \{1, 2, \cdots, T\}$. The

[8]There is only one block each pile in the probe template, so no need for the layer subscript.

identity can be retieved by the parameter t following Equation 4.7.

$$\text{vote}(\mathcal{P}B_{r,c}) = arg \max_{t\in\{1,2,\cdots,T\}} (\text{Sim}\{\mathcal{G}P^t_{r,c}, \mathcal{P}B_{r,c}\}) \tag{4.7}$$

From the perspective of algorithm design, to match a probe $\mathcal{P}$ against a gallery set of T images, LBP-DLMA involves two stages: an offline process that prepares the gallery templates, one template for each image, and an online process that prepares the probe template and performs the comparison, as shown in Table 4.1 and Table-4.2.

4.4 Another Version of LBP-DLMA: LBP-DTMA

The original motivation of adopting regional voting scheme is to conquer the registration difficulty caused by the non-rigidity of facial features. Regional voting works on the hard combination (local decisionsions within each block generate the final decision by majority voting) of the blocks. It outperforms the soft combination (similarity values obtained in all blocks are accumluted to generate the final simlairty value) in general[14]. However, to a specific application, such like face recognition, we still see some ground for adopting soft combination.

An example is shown in Figure 4.4 and Figure 4.5. Figure 4.4 shows the best block similarities of each gallery face $\mathcal{G}^t$. By applying LBP-DLMA, each block in $\mathcal{P}$ gets a vote for the identity whose block similarity is highest among all. The voting result is shown is Figure 4.5(a) with the final identity decision goes to the identity of $\mathcal{G}^1$ who gains 7 votes out of 12. However, if we apply the soft combination, summing up all block similarities of $\mathcal{G}^t$ to obtain their global similarities respectively to find the best matched $\mathcal{G}^{opt}$ the result would be overturned. Figure 4.5(b) shows that final identity decision goes to the identity of $\mathcal{G}^2$ because it has the highest similarity

Table 4.1: LBP Displacement-based Local Matching Approach - Off-Line

---

Perameters Chosen: Number of Piles in Each Image $K \times L$ (K piles per column, L piles per row), Shifting Value s, Number of Windows per Block $w_c \times w_l$ (w_c piles per column, w_l piles per row).

A. Off-Line Gallery Image LBP Displacement Description Construction:

Require: a gallery of face images; the size of the gallery is T.

For each image $\mathcal{G}$

1. Obtain the pixel label map by calculating the LBP pattern of each pixel (Note: The label map is slightly smaller than the original gallery image since the pixels on the boundaries may not have a label.) Assume the smaller size is $(m + 2s) \times (n + 2s)$.

2. For $i = 0$ to $2s$

   2.1. For $j = 0$ to $2s$

      2.1.1. Remove i, $2s - i$, j and $2s - j$ pixels from the leftmost, rightmost, topmost and bottommost boundaries of the label map to obtain a layer. (Note: as a total, there are $(2s + 1)^2$ layers.)

      2.1.2. Partition this layer into $K \times L$ blocks; partition each block into $w_c \times w_l$ windows, where we obtain the LBP label statistics (histogram of pixel labels); then concatenate the LBP label statistics of all windows in each block into a block level LBP description.

3. Obtain the LBP displacement description of the gallery image by piling up the corresponding block level LBP descriptions into each pile.

Table 4.2: LBP Displacement-based Local Matching Approach - On-Line

B. On-Line Face Recognition:
Require: $\mathcal{P}$ is an $(m+2s) \times (n+2s)$ sized probe image.

B-1 Obtain the LBP displacement description for $\mathcal{P}$ as follows:

1. Obtain the pixel label map by calculating the LBP pattern of each pixel.
2. Remove s pixels from all four sides of the label map.
3. Partition the label map into $K \times L$ blocks.
4. Partition each block into $w_c \times w_l$ windows, where we obtain window level LBP statistics, then concatenate the window level LBP statistics into a block level LBP description; each block LBP description constructs an LBP displacement pile; the set of all LBP piles is the LBP displacement description.

B-2 Do classification as follows:

1. Set vote counters $V_t = 0$ for all $t \in \{1, 2, \cdots, T\}$.
2. For r=1 to K
   2.1. For c=1 to L
      2.1.1. For $\mathcal{GP}^t_{r,c}$, where $t \in \{1, 2, \ldots, T\}$
         2.1.1.1 Calculate $\text{Sim}(\mathcal{GP}^t_{r,c}, \mathcal{PP}_{r,c})$, according to Equation 4.5.
      2.1.2. Find image index $I = \arg\max_{t \in \{1,2,\cdots,T\}} \text{Sim}(\mathcal{GP}^t_{r,c}, \mathcal{PP}_{r,c})$.
      2.1.3. Increase V_I by 1.
3. Classify the image as the identity of image g_J in the gallery set, where $J = \arg \max_{t \in \{1,2,\cdots,T\}} V_t$.

with $\mathcal{P}$ as a whole.

It would be an endless discussion that in this example, which face, $\mathcal{G}^1$ or $\mathcal{G}^2$, look most like $\mathcal{P}$. Regardless of the answer itself, we believe such a discussion has its theoretical contributions. This discussion gives us the insight to generate soft combination and thus come to the second version of our algorithm, the direct template matching approach, we name LBP-DTMA.

The main idea of LBP-DTMA is finding the best aligned face for every gallery image $\mathcal{G}^t$ and calculate the global similarities of all best aligned faces. The one that has highest similarity claims the identity of $\mathcal{P}$. LBP-DTMA works as follows[9]:

Given a gallery set of size T and a probe image $\mathcal{P}$, we obtain the template descriptions for each $\mathcal{G}^t$ and $\mathcal{P}$ as in LBP-DLMA.

Based on previous observations, the best aligned face can be found locally following Equation 4.8. For each $\mathcal{G}^t$, find its best aligned candidate face that satisfies Equation 4.5 and calculate its similarity to $\mathcal{P}$. This similarity is denoted by $Sim(\mathcal{G}^t, \mathcal{P})$ and taken as the similarity between $\mathcal{G}^t$ and $\mathcal{P}$. $\mathcal{P}$ shares identity of the one that has the highest similarity among all $\mathcal{G}^t$, as in Equation 4.9.

$$\begin{aligned} \mathrm{Sim}(\mathcal{G}, \mathcal{P}) &= \max_{1 \le l \le (2s+1)^2} \sum_{r,c} \mathrm{Sim}(\mathcal{G}B_{r,c,l}, \mathcal{P}B_{r,c}) \\ &= \sum_{r,c} \max_{1 \le l \le (2s+1)^2} \mathrm{Sim}(\mathcal{G}B_{r,c,l}, \mathcal{P}B_{r,c}) \end{aligned} \tag{4.8}$$

$$\mathrm{ID}(\mathcal{P}) = arg \max_{t \in \{1,2,\cdots,T\}} (\mathrm{Sim}(\mathcal{G}^t, \mathcal{P})) \tag{4.9}$$

Algorithms are shown in Table 4.1 and Table 4.3.

[9]Any undefined symbols/terms we use in LBP-DTMA are applied from LBP-DLMA and all symbols we use in LBP-DTMA are consistent with the symbols from LBP-DLMA if not otherwisely defined.

Table 4.3: LBP Displacement Template Matching Approach-On-Line

ON-LINE FACE RECOGNITION:
Require: $\mathcal{P}$ is a probe image.

1. Obtain the description template for $\mathcal{P}$ as follows:

   1.1. Obtain the pixel label map by calculating the LBP pattern of each pixel as in approach $\mathcal{A}$.

   1.2. Remove s pixels from all four sides of the LBP map.

   1.3. Partition the label map into $K \times L$ blocks;

   1.4. Partition each block into $w_l \times w_c$ windows, where we obtain window level descriptions (histogram of pixel labels), then concatenate the window level descriptions into a block level description.

2. Do classification as follows:

   2.1 For the description templates of each gallery image $\mathcal{G}^t$ (where $t \in \{1, 2, \cdots, T\}$)

      2.1.1 For each pile $\mathcal{G}P_{r,c}$ in the probe template

         2.1.1.1 For each $\mathcal{G}B^t_{r,c,l}$, in the pile $\mathcal{G}P^t_{r,c}$

            2.1.1.1.1 Calculate $\mathrm{Sim}(\mathcal{G}B^t_{r,c,l}, \mathcal{P}B_{r,c})$ according to Equation (1), (2), (3) or (4) (as the similarity formula chosen)

   2.2 Let $S^t = \sum_{r,c} \max_l(\mathrm{Sim}(\mathcal{G}B^t_{r,c,l}, \mathcal{P}B_{r,c})$.

   2.3 Classify the image as the identity of image $\mathcal{G}^I$ in gallery, where $I = \arg \max_{t \in 1,2,\cdots,T} S^t$.

57	58	12	58	57	37	37	37	43	43	23	25	25	21
58	6	58	57	38	38	37	29	48	42	58	25	25	28
20	34	41	38	37	38	1	58	1	1	0	11	25	25
58	57	44	38	3	3	5	15	14	14	41	58	25	25
57	44	37	3	58	33	58	27	57	28	3	0	58	25
57	38	2	35	32	39	58	26	25	14	35	33	0	25
31	37	58	57	45	44	45	23	54	54	57	45	58	18
39	38	37	43	23	48	43	42	52	43	43	30	22	26
38	38	58	0	19	58	37	22	58	30	30	39	22	18
31	38	5	13	35	33	3	0	14	34	6	3	0	19
32	31	4	19	57	58	15	15	8	57	6	10	8	19
32	32	58	58	0	14	56	57	6	0	34	40	8	19
33	32	22	26	35	58	55	44	5	57	40	0	58	19
57	32	0	58	57	58	0	11	57	57	38	4	13	19
53	33	6	0	58	50	35	58	52	37	16	13	19	57
58	57	33	32	0	51	57	46	37	0	13	19	57	58
25	53	40	32	6	0	58	1	0	13	19	26	53	58
25	52	50	32	39	10	10	9	14	27	26	18	53	39

(a) Original LBP Map of Size 18×14

6	58	57	38	38	37	29	48	42	58	25	25
34	41	38	37	38	37	1	58	1	0	11	25
57	44	38	3	3	5	15	14	14	41	58	25
44	37	3	58	58	58	27	57	28	3	0	58
38	2	35	32	39	58	26	25	14	35	33	0
37	58	57	45	44	45	23	54	54	57	45	58
38	37	43	23	48	43	42	52	43	43	30	22
38	58	0	19	58	37	22	58	30	30	39	22
38	5	13	35	33	3	0	14	34	6	3	0
31	4	19	57	58	15	15	8	57	6	10	8
32	58	58	0	14	56	57	6	0	34	40	8
32	22	26	35	58	55	44	5	57	40	0	58
32	0	58	57	58	0	11	57	57	38	4	13
33	6	0	58	50	35	58	52	37	16	13	19
57	33	32	0	51	57	46	37	0	13	19	57
53	40	32	6	0	58	1	0	13	19	26	53

(b) 16×12 Sized LBP Map From (a)

Figure 4.1: LBP Map

57	37	37	37
38	38	37	29
37	38	1	58
3	3	5	15

(a) Block 1

37	37	37	43
38	37	29	48
38	1	58	1
3	5	15	14

(b) Block 2

37	37	43	43
37	29	48	42
1	58	1	1
5	15	14	14

(c) Block 3

38	38	37	29
37	38	1	58
3	3	5	15
58	33	58	27

(d) Block 4

38	37	29	48
38	1	58	1
3	5	15	14
33	58	27	57

(e) Block 5

37	29	48	42
1	58	1	1
5	15	14	14
58	27	57	28

(f) Block 6

37	38	1	58
3	3	5	15
58	33	58	27
32	39	58	26

(g) Block 7

38	1	58	1
3	5	15	14
33	58	27	57
39	58	26	25

(h) Block 8

1	58	1	1
5	15	14	14
58	27	57	28
58	26	25	14

(i) Block 9

Figure 4.2: A pile of LBP displacement blocks of the LBP map in Figure 4.1(a)

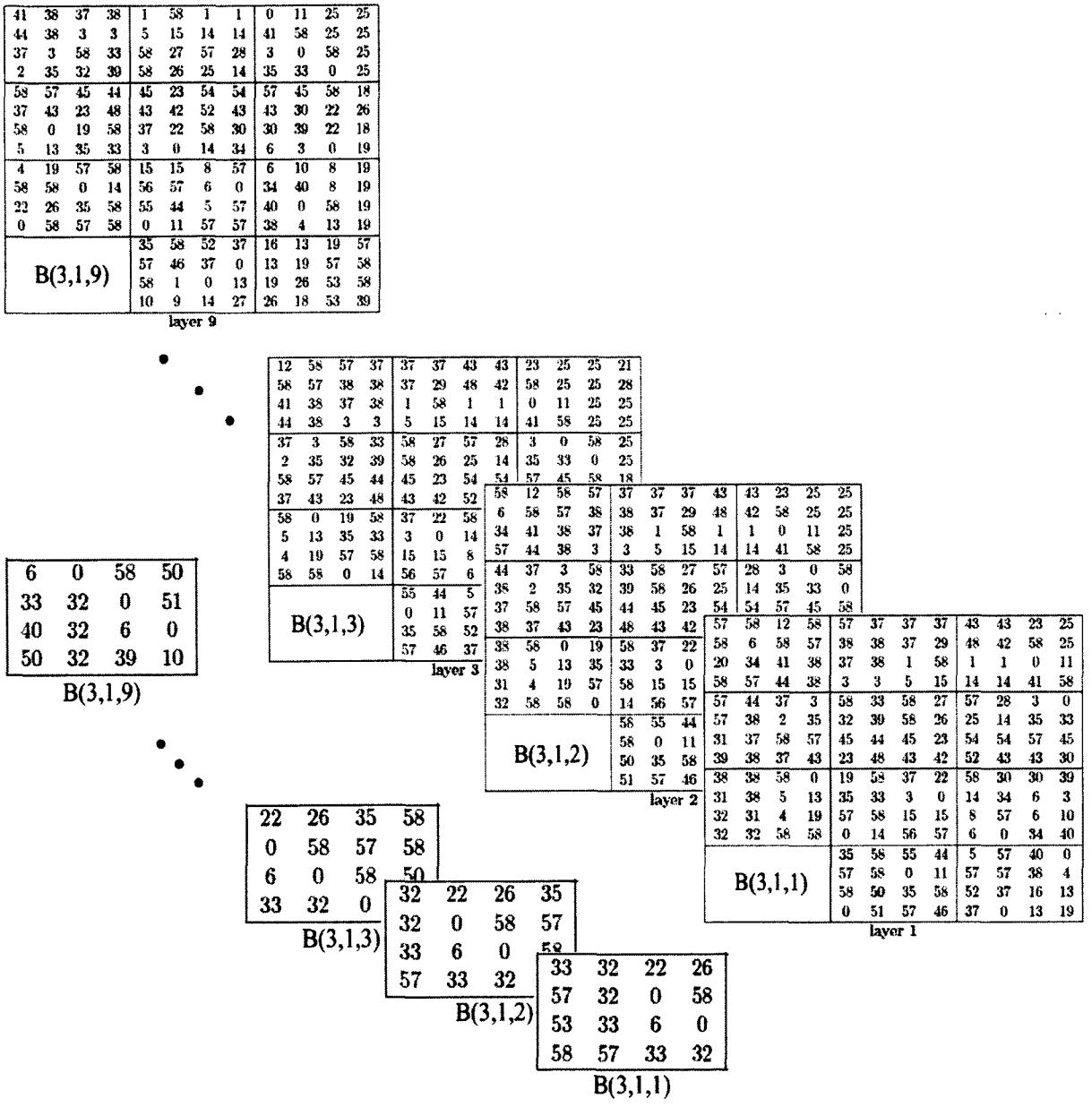


Figure 4.3: The LBP displacement description of the face in Figure 2.2 and an amplified pile $\mathcal{P}_{3,1}$

0.8	0.8	0.8
0.8	0.8	0.8
0.8	0.2	0.2
0.2	0.2	0.2

(a) Best Block Similarity for $\mathcal{G}^1$

0.7	0.7	0.7
0.7	0.7	0.7
0.7	0.7	0.7
0.7	0.7	0.7

(b) Best Block Similarity for $\mathcal{G}^2$

0.1	0.1	0.1
0.1	0.1	0.1
0.1	0.1	0.1
0.1	0.1	0.1

(c) Best Block Similarity for $\mathcal{G}^t$, $t \neq 1, 2$

Figure 4.4: Best block similarity for every gallery image in a gallery set compared with a probe image $\mathcal{P}$

$\mathcal{G}^1$	$\mathcal{G}^1$	$\mathcal{G}^1$
$\mathcal{G}^1$	$\mathcal{G}^1$	$\mathcal{G}^1$
$\mathcal{G}^1$	$\mathcal{G}^2$	$\mathcal{G}^2$
$\mathcal{G}^2$	$\mathcal{G}^2$	$\mathcal{G}^2$

(a) Block Results for $\mathcal{P}$ by Hard Combination: $ID(\mathcal{P}) = ID(\mathcal{G}^1)$

$Sim(\mathcal{G}^1, \mathcal{P}) = 7.6$

$Sim(\mathcal{G}^2, \mathcal{P}) = 7.9$

$Sim(\mathcal{G}^t, \mathcal{P}) = 1.2$

(b) Block Results for $\mathcal{P}$ by Soft Combination: $ID(\mathcal{P}) = ID(\mathcal{G}^2)$

Figure 4.5: Comparison results of local voting and template

Chapter 5

Experiments

We carry out experiments on FERET [34], "Faces in the Wild" (LFW) [28] and FRGC [35]. The usage of large, well developed databases avoids the bias from the images[1][34] and experimental results following the restrictions of the datasets are compared on the same platform to provide a more convincing evaluation for the algorithms.

LBP descriptor involves a few parameters. In all our experiments, as suggested in [3], we set the parameters as in Table 5.1. To further reduce the number of LBP displacement blocks in each pile, we restrict the relative offset by restricting $|3-i|+|3-j| \leq 4$. It is understandable that, in a practical system we may further improve the accuracies if we adjust these parameters on a "trial and error" basis though we do not include such a strategy in this work as we believe it is not necessary for research purposes[2].

[1] As Phillips et al mentioned in [34]: "Before the database FERET, a large number of papers reported outstanding recognition results usually > 95 percent correct recognition on limited-size database usually < 50 individuals. "

[2] "If you torture the data long enough, it will confess." —Ronald Coase.

Table 5.1: Parameters in our experiments

LBP Operator: $LBP^{u2}_{8,2}$	
radius of circle	R=2
number of sampling points	P=8
apply uniform pattern	yes
LBP-DLMA:	
number of blocks per LBP map	5×5 (K=5,L=5)
number of windows per block	7×7
margin cut from the LBP map	s=3
margins on top,bottom,left,right	$h, 6-h, j, 6-j > 0$;
other restrictions	$\|3-h\| + \|3-j\| \leq 4$

5.1 FERET

FERET database [34] is assembled to test and evaluate face recognition algorithm under standard tests and procedures. FERET consists of 14051 gray-scale images from 1199 individuals. The images vary in lighting conditions, facial expressions, pose azimuths, etc.. Subsets are presented with different task concerns.

We carry out experiments on FERET. Following the work in [3], five sets of FERET are used: the Fa gallery set that contains images of 1196 subjects, one image for each subject; the Fb probe set that contains 1195 face images of 1195 subjects as in Fa but with alternative facial expressions; the Fc probe set that contains 194 face images taken under different illumination conditions on the same day as their respective Fa matches; the Dup1 probe set that contains 722 face images taken anywhere between one minute and 1031 days after the corresponding images in Fa were taken; the Dup2 probe set being a subset of dup1 that contains 234 face images taken at least 18 months after the corresponding Fa images were taken. These five sets are designed for the study of algorithm performance against facial expressions(Fa, Fb), illuminations(Fa, Fc) and aging(Dup1, Dup2).

All faces are first normalized into standard size 150×130 (150 pixels per column, 130 pixels per row), where the distance between the centers of the two eyes is 56 pixels and the segment connecting centers of two eyes lies on the 53rd pixel below the top boundary. The standard 150×130 elliptical mask from FERET data collection is used to exclude non-face areas from the LBP maps, and a few pixels are removed from each side of the mask since the LBP map of an image is always smaller than the original image.

Following [3], permutation test with 95% confidence level is also carried out using the image list, list640.srt, in the CSU face identification evaluation system package [7]. list640.srt contains 4 images each for 160 subjects. 10000 permutations are tested, with each containing one image per subject in the gallery set and another in the probe set.

The results are shown in Table 5.2. The results of a few famous approaches are listed in the same table for comparison.

It is shown that LBP-DLMA not only improves the original LBP approach, but also achieves the performances at least comparable to the state of the art approaches.

It was explained in [48] that a preprocessing stage can significantly improve the performance of LBP approach. Therefore, we also do the experiments with the preprocessing as suggested in [48]. Results are shown in Table 5.3.

5.2 FRGC

We carry out the FRGC experiment 104 [35] of FRGC version 1, which is generally considered the most challenging in this FRGC V1 dataset. It requires recognizing 608 uncontrolled faces from 152 controlled gallery faces.

We normalize the face images into size 150×130 as we did for FERET exper-

Table 5.2: The recognition rates of the original LBP and weighted LBP, the LBP-DTMA, and LBP-DLMA for the FERET probe sets, the mean recognition rates of the Fb+Fc+Dup1, and results of permutation test with a 95% confidence level.

Method		Fb	Fc	Dup1	Dup2	Fb,Fc & Dup1	Permutation Test		
							lower	mean	upper
LBP, no weight [4]		93%	51%	61%	50%	78.20%	71%	76%	81%
LBP, weighted [4]		97%	79%	66%	64%	84.74%	76%	81%	85%
LBP-DLMA	Euclidean Distance	99.37%	93.60%	79.66%	75.56%	92.10%	84.92%	89.24%	93.31%
	Histogram intersection	99.39%	96.16%	82.52%	80.31%	93.32%	87.21%	91.22%	95.09%
	Chi square statistic	99.31%	96.20%	82.23%	80.53%	93.18%	87.34%	91.33%	95.18%
LBP-Template	Euclidean Distance	98.49%	90.21%	70.50%	61.11%	88.16%	78.13%	83.26%	88.13%
	Histogram intersection	98.91%	92.78%	76.04%	68.38%	90.53%	83.13%	87.88%	92.50%
	Chi square statistic	98.74%	91.24%	75.62%	65.81%	90.15%	83.13%	87.61%	91.88%

Table 5.3: The recognition rates of the LBP-DTMA, LBP-DLMA boosted by preprocessing schemes on the FERET probe sets, and a few known approaches.

Method		Fb	Fc	Dup1	Dup2
Preproceed LBP-DLMA	Euclidean Distance	99.29%	98.97%	85.37%	82.29%
	Histogram intersection	99.37%	99.48%	88.40%	85.89%
	Chi square statistic	99.37%	99.25%	88.71%	86.89%
preproceed LBP-DTMA	Euclidean Distance	98.49%	98.45%	84.07%	82.05%
	Histogram intersection	99.00%	98.97%	88.23%	86.75%
	Chi square statistic	99.00%	98.45%	88.23%	86.75%
LGBPHS[63]		98.0%	97.0%	74.0 %	71.0%
HGPP[61]		97.6%	98.9%	77.7 %	76.1%
SIS [32]		91.0%	90.0%	68.0 %	68.0%
Schwartz [39]		95.7%	99.0%	80.3 %	80.3%

iments. The results are shown in Figure 5.4. We also include the results of LBP-DLMA with a "preprocessing" stage, as suggested by Tan et al [48]. We can see that LBP-DLMA with and without preprocessing improve LBP with and without preprocessing significantly.

We should emphasize here that, our intension is to improve LBP approaches by using local matching scheme. It is not our intension to show that our approach is better than all possible approaches in all datasets. We understand some other approaches, such as [39], get better results for this experiment; we should add that those approaches actually use the settings more flexibly than we do —they use a training approach while we do not.

5.3 LFW

We have also carried out experiments on "Labeled Faces in the Wild" (LFW)[3] [28]. LFW is a database containing 13,233 face images of 5,479 individuals collected from

[3]The set is available via LFW official site http://vis-www.cs.umass.edu/lfw.

Table 5.4: Recognition rates of LBP-DLMA approaches on FRGC Experiment 104

LBP[26]	LBP DLMA		
	Euclidean	Histogram intersection	Chi square statistics
28.1%	34.38%	32.17%	33.23%
LBP	LBP DLMA with Preprocessing		
with Preprocessing[26]	Euclidean	Histogram intersection	Chi square statistics
58.1%	58.31%	67.47%	67.20%
LBP	LBP Template		
	Euclidean	Histogram intersection	Chi square statistics
28.1%	42.94%	47.37%	47.20%
LBP	LBP Template with Preprocessing		
with Preprocessing[26]	Euclidean	Histogram intersection	Chi square statistics
58.1%	74.01%	85.86%	86.18%

the web for the study of unconstrained face recognition. The faces were detected by the viola-Jones face detector and labeled by the name of individuals. 1,680 individuals in the database have two or more distinct photos. We test the performance of our approach on the 10 folds of view 2. All the face images were taken in unconstrained environments, exhibiting " 'natural' variability in pose, lighting, focus, resolution, facial expression, age, gender, race, accessories, make-up, occlusions, background, and photographic quality" [28]. In this task, given two face images, the goal is to decide whether two images are of the same person. This is a binary classification problem, with two possible outcomes: "same" or "different". LFW view 2 provides 10 folds of face sets where the sets of people in different folds are disjoint; when testing on one fold, the other nine folds can be used for training. Results of various approaches have been reported at LFW official site[4].

We use LFW-a version of images (the images aligned using a commercial face alignment software) [46]. The images are of size 250 × 250. We first crop them into images of size 90 × 78 (by removing 88 pixel margins from top, 72 from bottom,

[4]Note that most of the approaches reported were developed only for the specific binary classification task; our approach was not intended to be applicable only to this kind of tasks.

and 86 pixel margins from both left and right sides). Note that there were errors in the alignment of many images; we just keep them as they were (so some of the final cropped faces indeed are not correctly aligned).

In LBP-DLMA, since a "voting" is required in each pile, we need a few "reference faces" to find relative values. Here, we use a dummy set as "reference faces": for the experiments in the i-th fold, we use the first images (named "***..-0001.jpg") of the first 10 individuals in the $(i-1)$th fold (when $i-1=0$, we use the 10th fold) as the dummy set.

For a pair of images x and y, for each pile, we first obtain the similarity array between x and the set consists of y and dummy set, then obtain the similarity array between y and the set consists of x and the dummy set; the average to these two arrays are taken so as to make local decision according the this array.

Our results are shown in Figure 5.1 and Table 5.5.

Due to the fact that LBP-DLMA does not have a training process, our approach should be compared to other no-training approaches as suggested in LFW site; we also include the Receiver Operating Characteristic(ROC) curves of all these no-training approaches SD-MATCHES (L & R system with SIFT descriptors and MATCHES flavour), H-XS-40 (Histogram of LBP features with Chi Square similarity measure and 40 windows), GJD-BC-100 (Gabor Jets Descriptors with Borda Count measure and 100 reference images) and LARK representation without supervision [40], which are available in both LFW site and [30], in Figure 5.1 and Table 5.5. We can see that the LBP DLMA, regardless the similarity metrics that it uses, is significantly better than all other approaches.

For the alternate version LBP-DTMA, we can either use or not use dummy set. The results are shown as following:

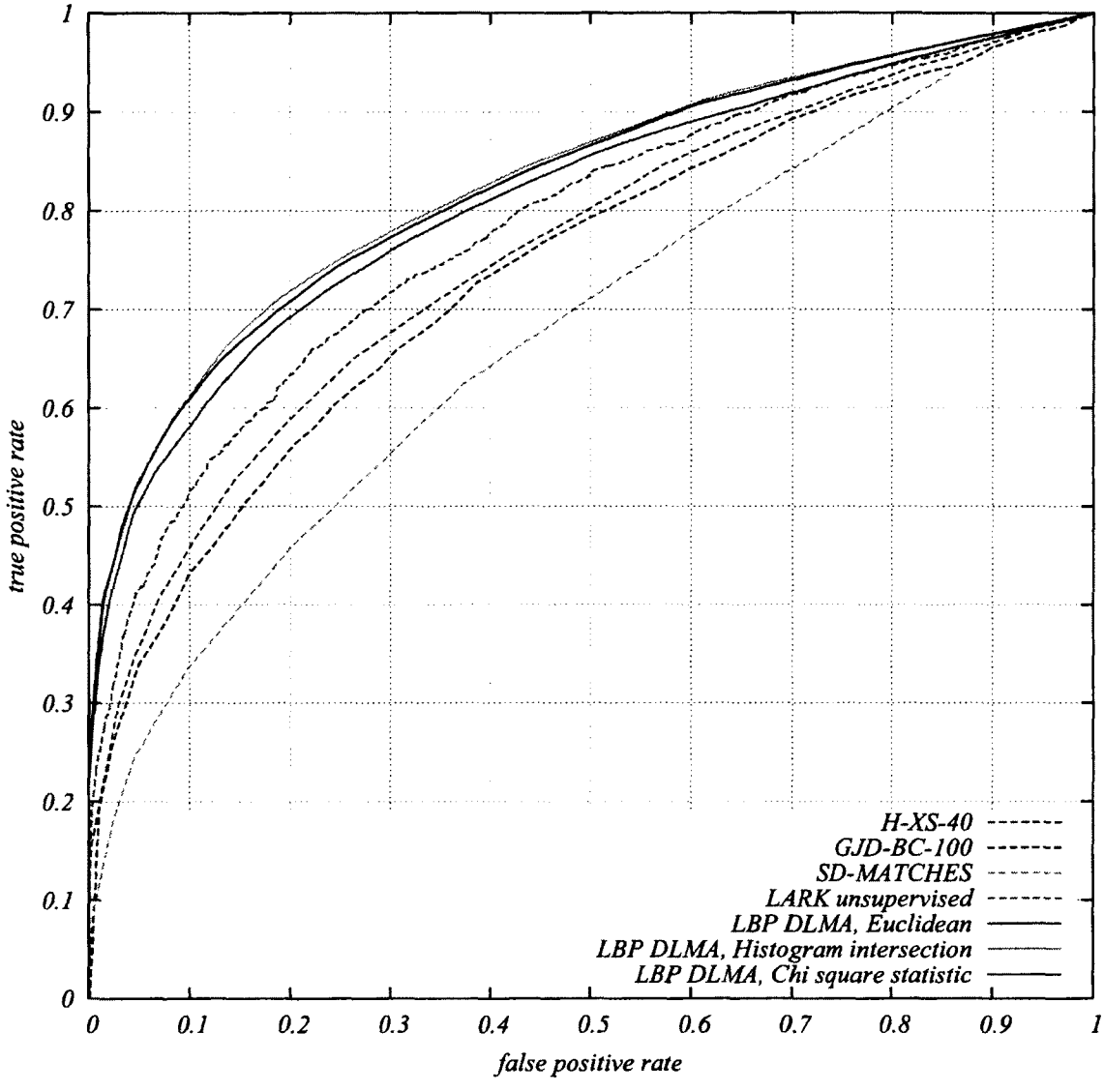


Figure 5.1: ROC curves over View 2 of LFW

Table 5.5: The accuracies of LBP-DLMA, LBP-DTMA and a few no-training approaches for LFW

Approach		Accuracy
SD-MATCHES		0.6410 ± 0.0062
H-XS-40		0.6945 ± 0.0048
GJD-BC-100		0.6847 ± 0.0065
LARK unsupervised		0.7223 ± 0.0049
LBP-DLMA	Euclidean	0.7517 ± 0.0122
	Histogram intersection	0.7648 ± 0.0186
	Chi square statistic	0.7622 ± 0.0206
LBP-DTMA	Euclidean	0.6905 ± 0.0235
	Histogram intersection	0.7428 ± 0.0144
	Chi square statistic	0.7417 ± 0.0143
LBP-DTMA, with Dummy Set	Euclidean	0.7352 ± 0.0180
	Histogram intersection	0.7633 ± 0.0152
	Chi square statistic	0.7613 ± 0.0172

Chapter 6

Extensibility

We expect that our approach can be applied to other descriptor approaches. Simply replacing LBP in Table 4.1 and Table 4.2 by any descriptor approach $\mathcal{A}$, we should able to generate $\mathcal{A}$-DLMA. We also expect that our approaches apply to low resolution images too.

6.1 Descriptors Other Than LBP

We test the extensionality of this displacement local matching approach on two variants of LBP: Three-Patch LBP (TPLBP) and Four-Patch LBP (FPLBP). Applying DLMA to TPLBP and FPLBP, we generate TPLBP-DLMA FPLBP-DLMA. Experiments of TPLBP-DLMA and FPLBP-DLMA are carried out on FERET datasets. For the parameters required for TPLBP and FPLBP, we use the default values of [57] as shown in Table 6.1:

The experimental results are shown in Table 6.2.

We can easily see that the performances of TPLBP-DLMA and FPLBP-DLMA are significantly better than TPLBP and FPLBP respectively.

Table 6.1: Parameters for TPLBP and FPLBP in our experiments

TPLBP Operator: $TPLBP_{2,8,3,5}$	
ring radius of circles	r=2
patch size	$3 \times 3, w = 3$
number of additional patches	S=8
distance between two apart patches	α=5
FPLBP Operator: $FPLBP_{4,5,8,3,1}$	
ring radii of two circles	$r_1 = 4$, $r_2 = 5$
patch size	3×3, w=3
number of additional patches	S=8
distance between two apart patches	α=1

6.2 Applications with Low Resolution Images

A mathematical assumption for local matching schemes being superior than global matching schemes is that the nation should be "large" enough [11, 14], although there is no fixed definition for "large". We now try to demonstrate that the LBP-DLMA also works for applications with small sized images.

We use [8]'s version of Yale and ORL face sets available via Cai's website `http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html`, where all faces are of standardized size 32×32. The Yale dataset contains the images of 15 subjects, each with 11 images captured with variations of lighting conditions and facial expressions such as normal, happy, sad, sleepy, surprised and wink). ORL dataset contains the images of 40 subjects, each with 10 images captured with variations of expressions and details such as open eyes, close eyes, smiling, no-smiling, w/o glasses. For any given k ($k = 2, 3, \cdots 8$), kTrain represents a split where k images per subject are chosen with labels for training, the rest are used for test. For fair comparison, 50 such random splits for each kTrain of both Yale and ORL are available via Cai's website.

We perform experiments on Yale and ORL with LBP approach and LBP-DLMA

Table 6.2: The recognition rates of original TPLBP, FPLBP, and TPLBP DLMA and FPLBP DLMA without / with Preprocessing [48] for the FERET probe sets, the mean recognition rate of the Fb+Fc+Dup1, and results of permutation test with a 95% confidence level.

Method		Fb	Fc	Dup1	Dup2	Fb,Fc & Dup1	Permutation Test		
							lower	mean	upper
TPLBP	Euclidean Distance	94.64%	74.23%	62.33%	55.98%	81.71%	68.13%	74.12%	80.00%
	Histogram intersection	96.44%	86.08%	74.65%	69.23%	88.04%	80.00%	85.06%	90.00%
	Chi square statistic	95.98%	86.08%	74.79%	69.66%	87.83%	79.38%	84.50%	89.38%
TPLBP DLMA	Euclidean Distance	99.26%	91.90%	75.97%	71.80%	90.62%	83.05%	87.51%	91.77%
	Histogram intersection	99.48%	95.15%	79.79%	75.7%	92.35%	85.68%	89.83%	93.91%
	Chi square statistic	99.38%	93.27%	78.83%	74.30%	91.79%	85.75%	89.90%	93.96%
Preprocessed TLBP DLMA	Euclidean Distance	98.88%	98.39%	77.56%	73.54%	91.54%	84.92%	89.27%	93.48%
	Histogram intersection	99.14%	98.23%	83.17%	81.98%	93.60%	87.87%	91.88%	95.68%
	Chi square statistic	99.15%	98.99%	82.31%	81.46%	93.38%	87.85%	91.85%	95.68%
FPLBP	Euclidean Distance	95.73%	69.59%	64.13%	54.70%	82.52%	72.50%	78.07%	83.13%
	Histogram intersection	96.65%	74.23%	67.45%	56.84%	84.60%	75.94%	81.19%	86.25%
	Chi square statistic	96.65%	74.23%	67.73%	56.41%	84.70%	75.63%	81.16%	86.25%
FPLBP DLMA	Euclidean Distance	98.89%	76.16%	68.68%	57.11%	86.47%	79.64%	84.32%	88.91%
	Histogram intersection	98.82%	81.09%	69.62%	60.98%	87.21%	80.84%	85.51%	90.09%
	Chi square statistic	99.04%	84.38%	70.56%	60.50%	87.95%	81.12%	85.78%	90.31%
Preprocessed FPLBP DLMA	Euclidean Distance	98.74%	98.24%	75.10%	69.65%	90.61%	84.01%	88.27%	92.45%
	Histogram intersection	99.00%	98.23%	76.96%	73.49%	91.39%	84.79%	89.07%	93.25%
	Chi square statistic	98.94%	98.22%	77.19%	73.08%	91.44%	85.05%	89.33%	93.49%

approach. For LBP approach, we let the window numbers per row (per column) to be 8, although the numbers between 7 and 9 seems to get very close accuracy. For LBP-DLMA, we let the number of blocks per row (per column) to be 4, and the number of windows in a block to be 3 per row (per column). Due the small size of the images, for LBP and for the LBP embedded within LBP-DLMA, we test on the radius of a circle to be both 2 and 1. For the sampling points distributed evenly on the circle, we keep it to be 8 as we did in Section "Experiments". The results are reported in Tables 6.3 and 6.4. We can easily find that LBP-DLMA can improve the accuracy of LBP approach regardless the parameters for generating the local binary pattern labels and regardless of the similarity measurements.

Table 6.3: Average Error Recognition Rates and Standard Deviations of LBP and LBP DLMA Algorithms, for Yale face set (32 × 32 pixels).

			2 Train	3 Train	4 Train	5 Train	6 Train	7 Train	8 Train
			Error±Std	Error±Std	Error±Std	Error±Std	Error±Std	Error±Std	Error±Std
Radius=2	LBP	Eucl.	43.16 ± 5.11	37.22 ± 3.87	35.52 ± 3.77	33.10 ± 3.41	30.59 ± 4.10	30.97 ± 3.39	30.40 ± 5.51
		Hist.	39.87 ± 4.95	34.88 ± 4.11	32.27 ± 3.07	29.99 ± 3.77	27.23 ± 4.43	27.48 ± 3.92	25.80 ± 5.33
		Chi	40.47 ± 5.22	35.58 ± 3.89	33.28 ± 3.34	31.20 ± 3.85	28.59 ± 4.47	28.23 ± 4.35	27.51 ± 5.55
	LBP	Eucl.	34.33 ± 4.94	28.09 ± 3.31	25.08 ± 2.86	23.06 ± 3.29	21.79 ± 3.67	20.75 ± 3.64	19.86 ± 4.27
		Hist	31.98 ± 4.56	25.68 ± 3.10	22.57 ± 2.16	20.72 ± 2.47	20.19 ± 3.36	18.88 ± 3.32	17.92 ± 4.07
	DLMA	Chi	32.85 ± 4.55	26.45 ± 3.10	23.98 ± 2.60	22.36 ± 2.95	21.32 ± 3.18	20.10 ± 4.06	19.18 ± 4.20
Radius=1	LBP	Eucl.	41.51 ± 5.69	35.41 ± 3.74	32.98 ± 3.77	31.22 ± 3.57	28.81 ± 3.57	29.33 ± 4.34	27.78 ± 5.93
		Hist.	36.64 ± 4.79	31.39 ± 3.89	29.69 ± 2.95	26.44 ± 3.72	24.39 ± 3.59	23.87 ± 3.82	22.11 ± 5.05
		Chi	37.50 ± 4.75	31.97 ± 3.70	30.15 ± 3.10	26.71 ± 4.06	24.32 ± 3.58	23.30 ± 4.01	21.29 ± 4.95
	LBP	Eucl.	31.08 ± 4.69	24.22 ± 3.07	21.23 ± 2.60	18.43 ± 3.27	17.46 ± 2.70	16.57 ± 3.41	15.09 ± 4.05
		Hist.	28.52 ± 4.08	23.26 ± 2.42	19.70 ± 2.42	16.58 ± 3.39	16.49 ± 3.23	14.50 ± 3.30	13.81 ± 4.02
	DLMA	Chi	29.11 ± 4.22	23.60 ± 2.92	20.16 ± 3.05	16.92 ± 3.89	17.54 ± 3.00	15.61 ± 2.94	14.50 ± 4.37

Table 6.4: Average Error Recognition Rates and Standard Deviations of LBP and LBP DLMA Algorithms, for ORL face set (32 × 32 pixels).

			2 Train	3 Train	4 Train	5 Train	6 Train	7 Train	8 Train
			Error±Std	Error±Std	Error±Std	Error±Std	Error±Std	Error±Std	Error±Std
Radius=2	LBP	Eucl.	19.97 ± 2.86	12.75 ± 2.00	8.40 ± 1.84	5.98 ± 1.77	4.49 ± 1.84	3.33 ± 1.59	2.26 ± 1.79
		Hist.	17.51 ± 2.75	10.80 ± 2.10	6.34 ± 1.64	4.29 ± 1.45	2.98 ± 1.60	2.14 ± 1.36	1.91 ± 1.78
		Chi	17.14 ± 2.75	10.81 ± 2.14	6.47 ± 1.58	4.33 ± 1.42	3.03 ± 1.50	2.15 ± 1.33	1.68 ± 1.63
	LBP	Eucl.	17.06 ± 2.49	10.07 ± 2.34	6.08 ± 1.45	3.97 ± 1.21	3.03 ± 1.21	1.77 ± 1.27	1.27 ± 1.17
		Hist	15.54 ± 2.19	9.16 ± 2.04	5.26 ± 1.19	3.36 ± 1.41	2.56 ± 1.23	1.56 ± 1.18	0.83 ± 0.83
	DLMA	Chi	16.88 ± 2.42	10.01 ± 2.18	5.89 ± 1.44	4.06 ± 1.28	2.95 ± 1.36	1.92 ± 1.29	1.18 ± 1.08
Radius=1	LBP	Eucl.	20.03 ± 3.04	13.19 ± 2.26	9.02 ± 1.70	6.12 ± 1.83	4.68 ± 1.65	3.38 ± 1.68	2.40 ± 1.51
		Hist.	16.35 ± 2.77	10.08 ± 1.83	6.29 ± 1.61	4.27 ± 1.21	3.28 ± 1.46	2.42 ± 1.42	1.93 ± 1.30
		Chi	16.49 ± 2.71	10.12 ± 2.14	6.13 ± 1.58	4.22 ± 1.26	3.18 ± 1.49	2.25 ± 1.56	1.70 ± 1.24
	LBP	Eucl.	17.08 ± 2.14	10.48 ± 1.74	6.49 ± 1.48	4.46 ± 1.36	3.13 ± 1.51	2.12 ± 1.44	1.51 ± 1.38
		Hist.	14.87 ± 2.30	8.62 ± 1.93	4.90 ± 1.22	3.09 ± 1.21	2.20 ± 1.27	1.22 ± 1.00	0.79 ± 0.93
	DLMA	Chi	16.38 ± 2.41	9.79 ± 2.26	5.80 ± 1.22	3.96 ± 1.24	3.15 ± 1.54	1.89 ± 1.12	1.50 ± 1.25

Chapter 7

Conclusion and Discussion

We introduce an LBP displacement concept so that LBP can be embedded into a local matching framework. The integration of LBP and regional voting, named LBP-DLMA, significantly improves the performances of the original LBP. Experiments also show that our approach can also be applied to descriptor approaches other than LBP and low resolution images.

The LBP-DLMA adopts local voting scheme, where winner-take-all is applied to select one "winner" when a pile of a probe is matched with a pile of a gallery image and a final decision is based on the votes of the regions (the blocks in the experiments). The introduction of block represents three levels of deviations: fixed deviation on a window level, dynamic deviation on a regional level and multi-deviation on a global level.

A following question is whether we can replace the local voting by "soft-combination", where the similarities of corresponding LBP displacement piles are added up to form the similarity between the LBP displacement descriptions of a pair of faces. Indeed it was shown in [10] that the answer to this question is positive.

It may be interesting to investigate the adoption of more complex strategies, such

as the randomized decision trees [33] for constructing/representing LBP displacement pile, and the learning of a similarity metrics [27] for exploiting the similarity values or assessments of all LBP displacement piles of a pair of LBP displacement descriptions.

Future work could also be a framework that could apply this DLMP approach to 3D images even motion pictures. As our approach aims at a relatively better alignment, we also see some insight on applying this DLMA to other comparison-based research.

Bibliography

[1] Y. Adini, Y. Moses, and S. Ullman. Face recognition: the problem of compensating for changes in illumination direction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):721–732, Jul. 1997.

[2] H. Agarwal, N. Jain, and M. Kumar. Face recognition using principle component analysis, eigenface and neural network. *Signal Acquisition and Processing, ICSAP 2010. International Conference on*, pages 310–314, Feb. 2010.

[3] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *Proc. of 9th European Conf. on Computer Vision*, pages 469–481, May 7-13 2004.

[4] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037 –2041, Dec. 2006.

[5] M. S. Bartlett, Movellan, R. Javier, and T. J. Sejnowski. Face recognition by independent component analysis. *Neural Networks, IEEE Transactions on*, 13(6):1450–1464, 2002.

[6] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):711–720, 1997.

[7] J. R. Beveridge, K. She, B. Draper, and G. H. Givens. A nonparametric statistical comparison of principal component and linear discriminant subspaces for face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, page 535542, Dec. 2001.

[8] D. Cai, X. He, Y. Hu, J. Han, and T. Huang. Learning a spatially smooth subspace for face recognition. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Jun. 2007.

[9] L. Chen. Pairwise macropixel comparison can work at least as well as advanced holistic algorithms for face recognition. In *Proceedings of the British Machine Vision Conference*, pages 5.1–5.11. BMVA Press, 2010. doi:10.5244/C.24.5.

[10] L. Chen, L.Yan, Y. Liu, L. Gao, and X. Zhang. Displacement template with divide-&-conquer algorithm for significantly improving descriptor based face recognition approaches. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *Computer Vision ECCV 2012*, volume 7576 of *Lecture Notes in Computer Science*, pages 214–227. Springer Berlin Heidelberg, 2012.

[11] L. Chen and N. Tokuda. Regional voting versus national voting –stability of regional voting (extended abstract). In *Int. ICSC Symposium on Advances in Intelligent Data Analysis*, Rochester, New York, USA, Jun. 22-25 1999.

[12] L. Chen and N. Tokuda. Robustness of regional matching scheme over globe matching scheme. *Artificial Intelligence*, 144(1-2):213–232, 2003.

[13] L. Chen and N. Tokuda. Stability analysis of regional and national voting schemes by a continuous model. *IEEE Trans. Knowledge and Data Engineering*, 15(4):1037–1042, 2003.

[14] L. Chen and N. Tokuda. A general stability analysis on regional and national voting schemes against noise – why is an electoral college more stable than a direct popular election? *Artificial Intelligence*, 163(1):47–66, 2005.

[15] L. Chen and N. Tokuda. A unified framework for improving the accuracy of all holistic face identification algorithms –electoral college for human face identification by computing machinery. *Artificial Intelligence Review*, 33(1-2), 2010.

[16] L. Chen, N. Tokuda, and A. Nagai. Robustness of regional matching over global matching –experiments and applications to eigenface-based face recognition. In M. R. Syed and O. R. Baiocchi, editors, *Intelligent Multimedia, Computing and Communications Technologies and Applications of the Future (Proc. of 2001 Int. Conf. on Intelligent Multimedia and Distance Education, Fargo, North Dakota, USA, June 1-3, 2001)*, pages 38–47. John Wiley & Sons, Inc., New York, 2001.

[17] Liang Chen and Ling Yan. Block lbp displacement based local matching approach for human face recognition. In Jong-Il Park and Junmo Kim, editors, *Computer Vision - ACCV 2012 Workshops*, volume 7728 of *Lecture Notes in Computer Science*, pages 97–108. Springer Berlin Heidelberg, 2013.

[18] S. Chen and Y. Zhu. Subpattern-based principle component analysis. *Pattern Recognition*, 37(5):1081–1083, 2004.

[19] MetaData Company. Metadata company website. `http://www.metadata.com.mx/`, 2013.

[20] Visionics Company. Visionics FaceIt®technology. `http://www.visionics.com/`, 2013.

[21] K. Etemad and R. Chellappa. Discriminant analysis for recognition of human face images. *Journal of the Optical Society of America*, 14:1724–1733, 1997.

[22] R. Frischholz. The Face Detection Homepage. `http://http://www.facedetection.com/`, 2013.

[23] T. Grüter, M. Grüter, and C. C. Carbon. Neural and genetic foundations of face recognition and prosopagnosia. *J Neuropsychol*, 2(1):79–97, 2008.

[24] H. Z. Gu and S. Y. Lee. Integrating two-dimensional morphing and pose estimation for face recognition with pose variations. *Journal of Information Science and Engineering*, 2012.

[25] J. V. Haxby, E. A. Hoffman, and M. I. Gobbini. The distributed human neural system for face perception. *Trends Cognitive Science*, 4:223–233, 2000.

[26] J. Holappa, T. Ahonen, and M. Pietikainen. An optimized illumination normalization method for face recognition. In *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, pages 1–6, Sept. 29-Oct. 1 2008.

[27] C. Huang, S. Zhu, and K. Yu. Large scale strongly supervised ensemble metric learning, with applications to face verification and retrieval. Technical Report TR115, NEC, 2011.

[28] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, Oct. 2007.

[29] ICBC. ICBC face recognition software for security. `http://www.icbc.com/`

driver-licensing/your-privacy, 2013.

[30] R. Javier, R. Verschae, and M. Correa. Recognition of faces in unconstrained environments: A comparative study. *EURASIP Journal on Advances in Signal Processing*, 2009. Article ID 184617, 19 pages.

[31] N. Kanwisher, J. McDermott, and M. M. Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*, 115(1):4302–4311, 1992.

[32] J. Liu, S. Chen, Z. Zhou, and X. Tan. Single image subspace for face recognition. In *AFMG*, pages 205–219, 2007.

[33] E. Nowak and F. Jurie. Learning visual similarity measures for comparing never seen objects. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[34] P. Phillips, H. Moon, S. A. Rizvi, and P. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Analysis & Machine Ingelligence*, 22(10):1090–1104, 2000.

[35] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proc. of Computer Vision and Pattern Recognition*, volume I, pages 947–954. San Diego, Jun. 2005.

[36] J. Sadr, I. Jarudi, and P. Sinha. The role of eyebows in face recognition. *Perception*, 32:285–293, 2003.

[37] P. Sanguansat, W. Asdornwised, S. Jitapunkul, and S. Marukatat. Class-

specific subspace-based two-dimensional principal component analysis for face recognition. 2006.

[38] Z. Schultz. Facial recognition technology helps DMV prevent identity theft. *WMTV-News*, 2007.

[39] W. R. Schwartz, H. Guo, and L. S. Davis. A robust and scalable approach to face identification. In *European Conference on Computer Vision*, pages 476–489, 2010.

[40] H. J. Seo and P. Milanfar. Face verification using the lark representation. *IEEE Transactions on Information Forensics and Security*, 6:1275–1286, Dec. 2011.

[41] S. Shan, W. Gao, B. Cao, and D. Zhao. Illumination normalization for robust face recognition against varying lighting conditions. *Analysis and Modeling of Faces and Gestures AMFG 2003. IEEE International Workshop on*, pages 157–164, 2003.

[42] P. Sinha, B. Balas, Y. Otrovsky, and R. Russell. Face recognition by human: Nineteen results all computer vision researchers should know about. *Proceedings of The IEEE*, 94(11):1948–1962, 2006.

[43] L. Sirovish and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A Optics and Image Science*, 4(3):519–524, 1987.

[44] M. Pietikäinen T. Ojala. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, 1996.

[45] M. Pietikäinen T. Ojala. Histogram of gabor phase patterns (hgpp): A novel

object representation approach for face recognition. *IEEE Transactions on Image Processing*, 16:57–68, 2007.

[46] Y. Taigman, L. Wolf, and T. Hassner. Multiple one-shots for utilizing class label information. In *The British Machine Vision Conference (BMVC)*, London, Sep. 2009.

[47] K. Tan and S. Chen. Adaptively weighted sub-pattern PCA for face recognition. *Neurocomputing*, 64:505–511, 2005.

[48] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *AMFG*, pages 168–182, 2007.

[49] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[50] M. Turk and A. Pentland. Face recognition using eigenfaces. *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 586–591, 1991.

[51] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mohahi, and Y. Ma. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(2):372–386, 2012.

[52] R. Walker, M. Stokes, M. Socker, and M. Collins. A study of the face recognition ability of orthodontists and lay persons of different age groups. *Journal of Orthodontics*, 39(1):9–16, 2012.

[53] P. Wang, L. C. Tran, and Q. Ji. Improving face recognition by online image alignment. *ICPR Pattern Recognition,18th International Conference on*, 1:311–

314, 2006.

[54] Y. Welinder. A face tells more than a thousand posts: Developing face recognition privacy in social networks. *Harvard Journal of Law & Technology*, 26(1):165–239, 2012.

[55] J. B. Wilmer, L. Germine, C. Chabris, G. Chatterjee, M. Williams, E. Loken, K. Nakayama, and B. Duchaine. Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences of the United States of America*, 107(11):5238–5241, 2010.

[56] Business Wire. Mexico adopts visionics' faceit technology in permanent system for eliminating duplicate voter registrations, 2000.

[57] L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In *Real-Life Images workshop at the European Conference on Computer Vision (ECCV)*, Oct. 2008.

[58] H. Xiong, M. N. S. Swamy, and M. O. Ahmad. Two-dimensiional FLD for face recognition. *Pattern Recognition*, 38(7):1121–1124, Jul. 2005.

[59] J. Yang, D. Zhang, A. F. Frangi, and J. Yang. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26:131–137, 2004.

[60] A. W. Young, D. Hellawell, and D. C. Hay. Configuration information in face perception. *Perception*, 16:747–759, 1987.

[61] B. Zhang, S. Shan, X. Chen, and W. Gao. Histogram of gabor phase patterns (hgpp): A novel object representation approach for face recognition. *IEEE*

Transactions on Image Processing, 16:57–68, 2007.

[62] D. Zhang, M. Yang, and X. Feng. Sparse representation or collaborative representation: Which helps face recognition? *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 471–478, 2011.

[63] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang. Local gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. *Computer Vision, ICCV 2005. Tenth IEEE International Conference on*, 1:786–791, 2005.

[64] J. Zou, Q. Ji, and G. Nagy. A comparative study of local matching approach for face recognition. *IEEE Transactions on Image Processing*, 16(10):2617–2628, 2007.